BIOLOGICAL STANDARDIZATION

OXFORD MEDICAL PUBLICATIONS

BIOLOGICAL STANDARDIZATION

BY

J. H. BURN

Professor of Pharmacology in the University of Oxford

D. J. FINNEY

Lecturer in the Design and Analysis of Scientific Experiment in the University of Oxford

AND

L. G. GOODWIN

Member of the Staff of the Wellcome Laboratories of Tropical Medicine

SECOND EDITION

GEOFFREY CUMBERLEGE
OXFORD UNIVERSITY PRESS
LONDON NEW YORK TORONTO

Oxford University Press, Amen House, London E.C. 4
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS CAPE TOWN
Geoffrey Cumberlege, Publisher to the University

FIRST EDITION 1937
SECOND EDITION 1950
SECOND IMPRESSION 1952

PRINTED IN GREAT BRITAIN
AT THE UNIVERSITY PRESS, OXFORD
BY CHARLES BATEY, PRINTER TO THE UNIVERSITY

PREFACE TO SECOND EDITION

When the first edition of this book was published I had been occupied daily for eleven years with biological standardization, both in finding new methods and in carrying out routine testing. Since 1937, however, I have been working on other aspects of pharmacology, and it has been difficult to prepare this second edition. The statistical side of standardization has expanded so rapidly that I have not attempted to discuss it and I have been fortunate in persuading Dr. Finney to deal with it. In the hope of ensuring that his exposition should not be too technical for less mathematically minded readers, I asked Dr. Pamela Holton and also Dr. R. S. Stacey to make comments upon it. Dr. Finney very willingly tempered the argument to meet them. It should be made clear that the whole responsibility for Chapter III is Dr. Finney's, just as that for Chapter II is mine. These two chapters have been written quite independently, and it is therefore possible that there are inconsistencies between them. I hope that my chapter will enable any beginner to get on with standardization without waiting until he has learnt to swim in the water of Dr. Finney's chapter. But he must learn to swim as soon as he can.

In the field of chemotherapy, the contributions made by Mr. L. G. Goodwin to our knowledge of methods of testing have been so important that I felt the book would be greatly strengthened if he would write a section about it. Chapters XXI to XXV are therefore Mr. Goodwin's responsibility (with the exception of the account of the testing of anticestode agents).

The methods for atropine substitutes, quinidine substitutes, curare-like compounds, substances which reduce gastric acidity, local anaesthetics and analgesics are also methods of testing like those for amoebicides and antimalarial compounds, and it is reasonable to take the view that they should not be included under the cover of *Biological Standardization.* I was sufficiently concerned about this to decide, while the writing was in progress, to change the title of the book to 'Quantitative methods in pharmacology and chemotherapy'. However, because of the continuity which the book has with the original edition, after a

month or two I changed it back again, though I am fully aware there is ground for criticism.

Attempts to draw sharp distinctions between terms like biological standardization, bio-assay, and biological methods of testing generally fail. For example, it has been suggested that biological standardization should be reserved for substances like insulin where there is no qualitative difference between the unknown and the standard. On this definition, however, we would no longer speak of the biological standardization of digitalis, although this was the first drug for which standardization was ever proposed!

The chapter on digitalis has been retained in this edition in its original form, though work has been done on this subject by many since 1937. Sir Thomas Lewis, the well-known cardiologist, always expressed the view that the standardization of digitalis was unnecessary. Speaking in the time before 1932, when tinctures of digitalis in Great Britain were not standardized, he said that the pharmacist at University College Hospital always informed him when a new batch of tincture was being sent to his wards, but that he had never been able to observe any variation in activity when he gave a new tincture to his patients. Lewis was a most careful and accurate observer, and his view that digitalis standardization was unnecessary requires consideration. He himself always urged that further work should be done to obtain crystalline active principles, and when digoxin was isolated he used it constantly. The particular difficulty offered by digitalis is the variation in the amount of different active principles in different batches of leaf, which means that different methods of standardization give different results. Since it has proved impossible to compare the potency of different samples in man with any accuracy, we do not know which animal method gives the right answer. In these circumstances it is surely right to choose the simplest method which takes the least time and requires animals or birds of which there is a plentiful supply.

The chapter on Vitamin D also remains unchanged from the first edition. I intended to omit it as the chapters on vitamins A, B, and C have been omitted, but it has been included because tests for vitamin D are still performed, and because Dr. Finney was using observations from it to illustrate statistical processes.

I am indebted to Professor J. H. Gaddum for notes he gave me on methods now used for the anterior pituitary hormones and the sex hormones; these notes saved much labour, for they pointed out which methods were most useful both from the point of view of accuracy and of simplicity.

I am conscious that many readers will feel that methods have been inexcusably omitted. Thus there is no reference to the standardization of heparin. My defence is that I have restricted the contents to methods of which I have some degree of first-hand knowledge. I have been less successful in this than in the original edition, but that has been the guide.

We wish to thank those who have allowed us to reproduce figures and tables; in particular Dr. E. C. Wood for the data used in Table XI, and Professor R. A. Fisher, Dr. F. Yates, Messrs. Oliver and Boyd, and the Cambridge University Press for some of the tables used as appendixes. We also thank Miss Elizabeth St. John for preparing the index (with the exception of that for Dr. Finney's chapter, which he prepared himself). Finally, I must call attention to the remarkable care which the Oxford University Press take in preparing the text, even to the point of recalculating means of columns of figures! It is a great pleasure to co-operate with them.

J. H. BURN

CONTENTS

CHAPTER I

UNITS OF MEASUREMENT

EVERY pharmacologist should visit Winchester. If he is also a student of *Realpolitik*, he can walk through the cloisters of the first English public school, founded by William of Wykeham in 1387, and watch the boys playing cricket in the meadow by the Itchen. If he is romantic he can look at the Norman transepts in the cathedral, or at the Round Table of King Arthur and his knights, where Tristan sat. And if he is just a pharmacologist, he can visit the Westgate where there are some of the old weights and measures. For Winchester was once the capital of England, though little indication of this remains in English life to-day except the large bottles known as 'Winchester quarts'. The things, however, of special interest in the tower of the Westgate are the examples of the standard yard. These are curious because they are not all of the same length, and to some, pieces have been added. It appears that the yard used to be the length of the king's arm, and therefore when a new king came to the throne the standard was changed.

When the names of the early units of length and measure are recalled, the foot, the digit, the palm, and the handful, it becomes evident that the standards were based on biological properties. These would be unsatisfactory to-day, since the length of a man's arm, or the width of his palm, whether he be king or not, is variable, and of little value as a standard for measurements requiring precision. But while pharmacologists would be the first to support the variability of the king's arm, they are slow to admit that all biological properties share the same variability. Every attempt which has hitherto been made to measure a substance by its physiological effect has been accompanied by the definition of a unit in terms of a biological property, and all these definitions imply that biological properties are not variable. The first attempt to measure the potency of digitalis was made by taking as a unit the amount required to kill 1 gramme of frog; that was known as the frog unit. Even to-day the activity of cardiac glycosides is sometimes expressed in frog units. How inconstant the frog unit is

may be seen from experiments in which the same sample of digitalis leaf has been examined at different times of the year. The figures in Table I are the values for the international standard digitalis powder in frog units; they range from 1,310 to 3,300 units per gramme. The first attempt to measure insulin was to define a unit as the amount required to produce convulsions in a rabbit; it was later modified as the amount required to reduce the blood-sugar of a fasting rabbit weighing 2 kg. to 0·045 per cent. All those who have worked with insulin know that this amount is far from constant, and the unit defined in this way was abandoned in 1925 by common consent.

TABLE I. *The activity of the standard digitalis powder determined at different times and expressed in frog units*

1935	21 September	1,670 frog units per gm. powder
	27 September	1,510 ,, ,,
	5 October	1,850 ,, ,,
	13 November	2,270 ,, ,,
	23 December	2,600 ,, ,,
1936	25 January	2,940 ,, ,,
	1 February	3,300 ,, ,,
	12 February	2,630 ,, ,,
	13 March	2,500 ,, ,,
	20 March	2,170 ,, ,,
	9 April	1,850 ,, ,,
	21 April	2,380 ,, ,,
	7 May	2,270 ,, ,,
	14 May	2,170 ,, ,,
	21 May	2,000 ,, ,,
	3 June	1,850 ,, ,,
	10 June	2,500 ,, ,,
	25 June	2,170 ,, ,,
	15 July	1,390 ,, ,,
	17 July	1,310 ,, ,,

A steady stream of experience that units defined in terms of physiological effects are very variable has had no effect on the introduction of similar new units. There have been the cat and pigeon units for digitalis, the rat and mouse units for the oestrus-producing hormone, the rabbit unit for the hormone of the corpus luteum, rat and guinea-pig units for different activities of the anterior lobe of the pituitary gland, dog units for cortical extract, and even in the year 1936 a mouse unit for the adrenotrophic hormone.

All these units are, like the length of the king's arm, very variable. They vary in the first place because of the variation of animals treated in exactly the same way in one laboratory; they vary in the second place because of the variation in the mean response of stocks of animals in different laboratories. Then these units depend on carrying out a test according to certain details of technique; but pharmacologists are never willing to copy each other's technique in all details, and, therefore, a third source of variation enters. The Pharisees were upbraided because they strained at a gnat but swallowed a camel; the pharmacologists to-day strain at the king's arm, but they swallow the frog, rat, and mouse, not to mention the guinea-pig and the pigeon.

Comparative methods. In contrast to this failure to make accurate measurements in terms of animal reactions, much greater success has attended investigations in which different workers have been asked to compare different preparations. If two preparations *A* and *B* have the same specific action, then it is usually easy to decide which is the stronger, and different workers usually obtain the same result, although their methods are dissimilar in important details. If two preparations of insulin are compared in different laboratories, the ratio of the potency of the two is usually found to be the same, although the one laboratory uses the rabbit blood-sugar method, while the second laboratory uses the mouse-convulsion method. Thus, when the object of the experiment is to compare two preparations so as to state their relative potency, details of technique become of much less importance. It is not true to state that these details cease to have importance, for sometimes they continue to play a part. But we can escape from strict uniformity in methods by turning all estimations into a comparison of two preparations.

Standard Preparations

The introduction of standards in the form of dry and stable preparations makes comparative methods possible, for every estimation of an unknown sample becomes a comparison with a standard to see how much stronger or how much weaker it is. This conception of a standard is due to Ehrlich, who introduced it to make easier the standardization of diphtheria antitoxin.

Ehrlich did not define the unit of antitoxin as the amount required to give a certain degree of protection to a guinea-pig, but instead he defined it as the amount of activity present in a given weight of one sample of dried antitoxin kept in his laboratory in Frankfurt. All estimations of unknown samples thereafter became a comparison with Ehrlich's standard antitoxin to discover what amount of the unknown sample contained the same activity as a given weight of Ehrlich's sample.

The first dry standard preparation for a substance other than a serum was the digitalis standard proposed and prepared by Magnus. If one worker states the potency of digitalis in frog units, and another in cat units, then since the frog unit and the cat unit are both variable quantities, it is impossible to compare figures in these two systems of units. If, on the other hand, the international standard digitalis powder be used, then the comparison of an unknown leaf with this standard powder can be made either by the frog method or the cat method, and the expression of the potency in units has no relation to the animal species.

There are now 39 of these standards kept either in the Serum Institute in Copenhagen, or in the National Institute for Medical Research in London; a list of these is given in Appendix I. As examples, we may take the standards for diphtheria antitoxin, for pituitary (posterior lobe) extract, and for vitamin A. The standard for diphtheria antitoxin is a quantity of dry diphtheria antitoxin collected in ampoules which were kept *in vacuo* over phosphorus pentoxide to ensure uniform dryness, and which were then sealed. The standard is kept in the Serum Institute in Copenhagen. The unit is the specific antitoxic activity contained in 0·0628 mg. of this powder. The standard for pituitary (posterior lobe) extract is a quantity of powder prepared from fresh posterior lobe material by treatment with acetone and subsequent drying over phosphorus pentoxide. It is similarly distributed in ampoules and stored at 0° C. The unit is the specific activity contained in 0·5 mg. of this powder. The standard for vitamin A is a quantity of crystalline β-carotene distributed in ampoules and stored at 0° C. The unit is the specific activity contained in 0·0006 mg. of this crystalline material. The last two standards are kept in the National Institute for Medical Research, London.

The adoption of stable standards of this kind brings the estimation of biological properties into the same position as the measurement of length and weight. The international biological standards in Copenhagen and in London are essentially similar to the standard metre and the standard kilogram.

Difficulties presented by stable standards. The theory underlying the adoption of stable standards is, as already explained, that when a biological estimation becomes a comparison between two preparations, the method used for the comparison ceases to be of importance. This is not always true, and several examples may be given where the comparison yields one answer by one method and a different answer by another method. When the British digitalis standard was prepared, to be used as required by the *British Pharmacopoeia*, 1932, it was carefully compared with the international standard digitalis powder in different laboratories by the cat method and the frog method. Different workers using the cat method found the British standard to contain 1·16 times as much activity per gramme as the international standard; workers using the frog method found the British standard to contain 1·4 times as much activity. Digitalis leaf contains several active principles; the relative toxicity of these differs for different species, and they are not present in the same proportion in different samples of leaf; hence, when a given sample is compared with the standard sample, a comparison is being made between preparations qualitatively unlike, and, therefore, of different relative toxicity for the cat and the frog. Hence the results of comparisons by different methods will not be the same.

A second example is provided by the active substances which possess the property of producing oestrus in ovariectomized animals. There are at the present time two standards for these substances, oestrone and oestradiol monobenzoate; the first standard prepared was oestrone, but it was found that when solutions containing esterified preparations of the hormone were compared with the standard oestrone, the relative potency varied according to the method of testing. A second standard was therefore adopted. There are several other oestrus-producing substances, the potency of which when compared with either of the two existing standards would vary according to the test applied. A third example is furnished by the anti-rachitic

vitamin; the standard for this is a solution of irradiated ergosterol, and the potency of cod-liver oil is usually determined by comparison with it. Generally rats are used for the test, but in theory it should be possible to use chickens instead. If, however, cod-liver oil is compared with irradiated ergosterol, its anti-rachitic potency appears much greater when the comparison is made with chickens than with rats. There is now reason to believe that the anti-rachitic substance in cod-liver oil is not identical with that in irradiated ergosterol.

In all these examples in which the result of the comparison with the standard varies according to the method used, the substances compared differ not only quantitatively but also qualitatively. In an ideal world there would be a standard for each substance possessing biological activity, and then the necessity for making a quantitative comparison between substances qualitatively unlike would not arise.

Although these disadvantages exist, and, with the identification of further substances having a common physiological action, may even increase in the future, on the whole the use of stable standards has proved a great advance in providing scientific workers and pharmaceutical firms with a uniform system of measurement.

A more widespread use of these standards is greatly to be desired in different countries. In Great Britain the standards have been prescribed in the *Pharmacopoeia* and biological methods depending on their use have been proposed. The introduction of the standards has had an important effect also in promoting scientific intercourse between workers of different nationalities.

Estimations when no standard exists. When a new active substance is described, there is always a period in which no international standard is available, and the question is sometimes raised whether during this period an animal unit can be used. This question is of importance to manufacturers who are making preparations for clinical use. Now the doctor is not helped when he is told that a preparation contains a certain number of mouse units or dog units per c.c.; what the doctor requires to know is that different preparations of one substance always possess the same potency; the manufacturer should therefore compare each new preparation directly with the previous one to ensure that

his product does not vary. This is the proper form of scientific control of potency. To be content with a determination of the number of mouse units per c.c., without comparison with the previous preparation, is to be content with a much less accurate form of standardization.

CHAPTER II

THE CLASSIFICATION OF METHODS

The Law of Variation. *The great progress which has been made in biological methods during the last twelve years has come very largely from the recognition of the fact of animal variation.* In 1922 the author was engaged in estimating the potency of digitalis by the Hatcher cat method, and results were obtained for one sample of leaf which varied in different cats; thus the lethal dose calculated in mg. leaf per kg. cat was 75 in one cat, 92 in another, and 126 in a third. The conclusion was drawn that the method was of no value, since the figures varied so much; there was a refusal to recognize the fact of variation. This refusal takes many different forms. Workers rightly spend very much time and labour in endeavouring to obtain animals of uniform sensitiveness by paying attention to breeding and diet and age and weight; some, however, believe that when they have done this, they can neglect the animal variation and carry out their experiments as though it did not exist; they refuse to recognize that the variation still persists even though it is diminished. Similarly, workers propose new methods of estimation which they claim are reliable because the animal response does not vary; these claims are like claims to have discovered perpetual motion.

The work of Magnus on digitalis. The first demonstration of the true position came from the laboratory of Magnus. In 1926 his pupil, de Lind van Wijngaarden, published the results of experiments on 573 cats in which he showed that there was a continuous variation in the sensitiveness of cats to digitalis. It was not that for the majority of cats the lethal dose of a given sample of digitalis leaf was the same, while for a minority the lethal dose was aberrant, but that the individual lethal doses were distributed over a wide range. When the individual lethal dose was taken as abscissa, and the frequency with which a particular lethal dose was met was taken as ordinate, de Lind van Wijngaarden obtained the chart shown in Fig. 1. The mean lethal dose was that which occurred with the greatest frequency, i.e. in the greatest percentage of cats; other lethal doses occurred

with a frequency declining steadily towards zero at points respectively 44 per cent. below and 56 per cent. above the mean lethal dose.

This work showed that the figure obtained for the lethal dose in one animal has little significance, and that the true lethal *dose is the mean value determined in large numbers of animals. When experiments are made with a given preparation on a suc-*

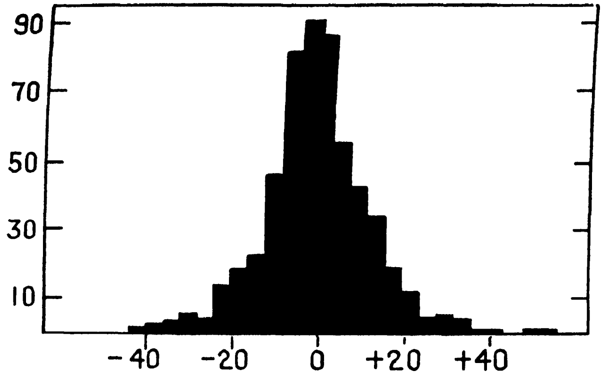


FIG. 1. The distribution around the mean value of the lethal doses of digitalis for different cats. The ordinates are numbers of cats. The abscissa 0 represents the mean value; the lethal doses vary from 44 per cent. below the mean to 56 per cent. above. (de Lind van Wijngaarden, *Arch. exp. Path. Pharmak.*, 1926, **113**, 40.)

cession of animals, the mean figure obtained at any point steadily approaches the true value as more animals are used.

Trevan and the lethal dose. The next advance was that made by Trevan, who in 1927 published an investigation of toxicity determinations of that kind in which the precise lethal dose cannot be found. In de Lind van Wijngaarden's work on digitalis the slow intravenous infusion of the drug made it possible to record the actual minimal lethal dose for every cat. When, however, toxicity is determined by a single injection, it is not possible to record more than the death or the survival of each animal, and the minimal dose is not discovered. When a descending series of doses is chosen, and each dose is injected into two or three animals, it has long been known that the results are irregular. Often animals survive which receive a dose high in the series, while animals die which receive a lower dose. Trevan investigated this anomaly and showed that when each

dose is injected into a sufficiently large group of animals, the irregular results disappear. There is a range of doses, represented in Fig. 2 by *OA*, which have no effect on any animal; the greatest dose in this range is the maximum tolerated dose; *similarly, there is a range of doses greater than OB which kill every animal; the least dose in this range may be called the certainly lethal dose.* Doses between *OA* and *OB* kill a percen-

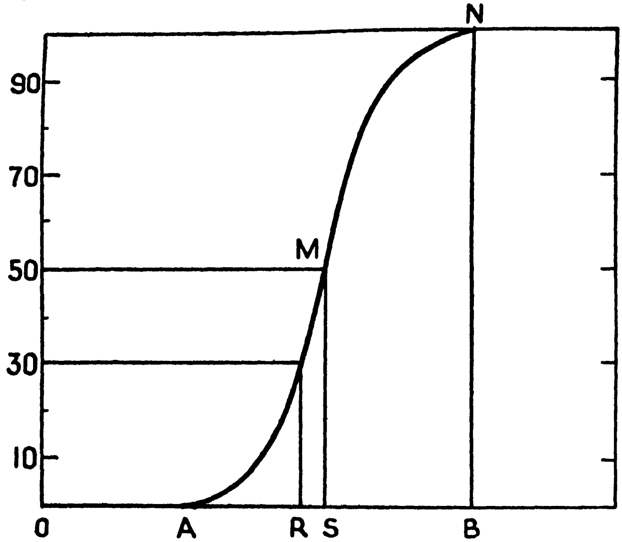


FIG. 2. The curve *AMN* shows the relation between percentage mortality (ordinates) and dose (abscissae). See text.

tage of animals which increases with the dose, and the graphic relation between percentage mortality and dose is a sigmoid or S-shaped curve like the curve *AMN* in Fig. 2. This curve rises most steeply in the neighbourhood of the dose causing 50 per cent. mortality. Perhaps the most surprising of the observations made by Trevan was the wide gap which separated the greatest dose which failed to kill any animal and the dose just great enough to kill every animal. The ratio of the latter dose to the former dose was determined for various substances. The figures are given in Table II.

These observations showed that the attempts usually made to determine a minimal lethal dose are exposed to great inaccuracy. Let us consider for a moment a definition of a minimal lethal dose as that dose which kills 2 out of 3 rats. If there are 99 rats

in a large cage and every one of the rats is injected with this dose, two-thirds will die and one-third will survive. But let us take them from the cage 3 at a time. Is it possible to be sure that 2 will die and 1 will survive out of the first 3 we inject? *Of course, there is no certainty. There are 33 rats in the cage which will survive the injection, and it is easily possible that 3*

TABLE II. *Ratio between the certainly lethal dose and the maximum tolerated dose (Trevan)*

	Species	*Ratio* LD 100/LD 0
Cocaine hydrochloride .	Mice	3·85
Digitalis . . .	Frogs	4·8
Echitamine . . .	Mice	5·5
Dysentery toxin . .	,,	20·0
Insulin (convulsions) .	,,	7·5

of these may be chosen. Only in a certain proportion of groups taken out of the cage will 2 rats die and 1 survive. In the other groups death will occur in 3 out of 3, or in 1 out of 3, or in 0 out of 3.

Qualitative Methods

Trevan's method of calculating LD 50. Trevan proposed that in tests of this kind the lethal dose should be determined which killed 50 per cent. of animals, and he introduced for this the term LD 50. As a rule the percentage mortality should be observed in a group of 30 animals. Trevan described the curve relating dose to percentage mortality as a characteristic for the species and the drug, which can be used to calculate the LD 50 from any observed percentage mortality. For example, let us suppose that a dose d of a substance T causes 30 per cent. mortality when injected into rats, and suppose also that the curve AMN in Fig. 2 is the characteristic for T when injected into rats. We note that the abscissa corresponding to 30 per cent. mortality is OR, and that the abscissa corresponding to 50 per cent. mortality is OS; from this it follows that the ratio of the dose causing 50 per cent. mortality to that causing 30 per cent. mortality is OS/OR. Now we have observed that a dose d caused 30 per cent. mortality; therefore LD 50 $= d \times OS/OR$.

Behrens's observations on strophanthin. The meaning of the

dose-mortality curve shown in Fig. 2 can best be understood from the work of Behrens (1929) who investigated the variation in the sensitiveness of frogs to *k*-strophanthin when administered by slow intravenous infusion. His apparatus enabled him to deliver a 0·3 per cent. solution at the rate of 0·01 c.c. in 30

TABLE III. *Behrens's figures for individual lethal doses of k-strophanthin for frogs*

Dose $\mu g.$	*No. of frogs killed by this dose*	*Percentage frequency*	*Integrated percentage frequency*
0·183–0·196	1	0·7	0·7
0·196–0·209	0	0·0	0·7
0·209–0·222	4	2·7	3·4
0·222–0·235	2	1·3	4·7
0·235–0·248	8	5·4	10·1
0·248–0·261	6	3·8	13·9
0·261–0·274	18	12·3	26·2
0·274–0·287	10	6·8	33·0
0·287–0·300	9	6·2	39·2
0·300–0·313	9	6·2	45·4
0·313–0·326	14	9·6	55·0
0·326–0·339	9	6·2	61·2
0·339–0·352	8	5·4	66·6
0·352–0·365	11	7·5	74·1
0·365–0·378	10	6·8	80·9
0·378–0·391	7	4·8	85·7
0·391–0·404	4	2·7	88·4
0·404–0·417	6	3·8	92·2
0·417–0·430	3	1·9	94·1
0·430–0·443	2	1·3	95·4
0·443–0·456	2	1·3	96·7
0·456–0·469	1	0·7	97·4
0·469–0·482	2	1·3	98·7

minutes. By this method Behrens determined the precise lethal dose for each frog. Observations made on 146 frogs showed that the average lethal dose was 0·326 $\mu g.$ per gm. but that some frogs were killed by 0·186 $\mu g.$ per gm., while others required as much as 0·483 $\mu g.$ per gm. Behrens divided the results into groups: thus he counted the number of frogs for which the lethal dose lay between 0·183 $\mu g.$ and 0·196 $\mu g.$ per gm., and again the number for which the lethal dose lay between 0·196 $\mu g.$ and 0·209 $\mu g.$ per gm., and so on. In this way he determined the 'frequency' with which frogs having a given sensitiveness were encountered. The figures he obtained are given in Table III. In the third

column of this table the frequency is expressed as a percentage; thus there were 8 frogs for which the lethal dose lay between 0·235 μg. and 0·248 μg.; since there were in all 146 frogs, the percentage frequency is 5·4. In the fourth column the percentage frequencies are integrated, that is to say, to each *percentage frequency is added the percentage frequency of frogs killed by lower doses.*

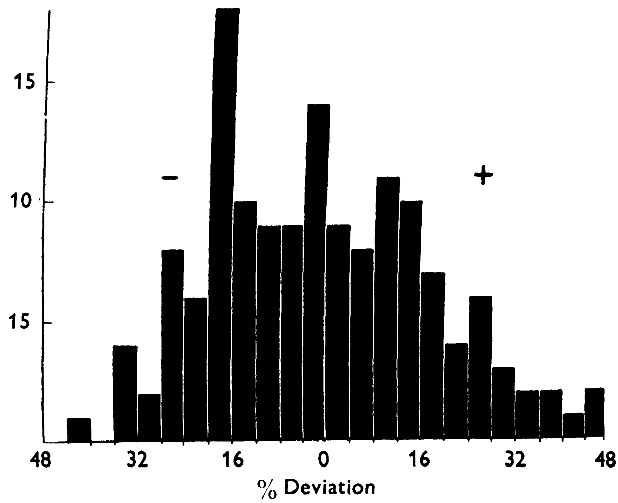


FIG. 3. Showing the distribution around the mean of the lethal doses of strophanthin for different frogs. The ordinates are numbers of frogs. (Behrens, *Arch. exp. Path. Pharmak.*, 1929, **140**, 237.) The deviations extend from 44 per cent. below the mean (shown on abscissa as 0) to 48 per cent. above. Compare Fig. 1.

From the results given in the first two columns of Table III Behrens constructed the chart shown in Fig. 3, in which the frequencies, that is to say, the number of frogs killed by a given range of doses, are plotted as ordinates and the lethal doses as abscissae. This chart is a frequency-distribution chart like that in Fig. 1, and it will be seen that it has a roughly similar outline.

The most important result of Behrens's experiments is obtained when the integrated frequencies given in the last column of Table III are plotted against the doses. The doses themselves are not used, but they have been recalculated so that the dose producing 50 per cent. of deaths has the value 0. The points are shown in Fig. 4, in which they fit very closely the dose-mortality

curve obtained experimentally by Trevan for digitalis. It is evident from the agreement that Trevan's dose-mortality curve is to be explained as an integrated frequency curve. Both Trevan and Behrens used *Rana temporaria*, but while Trevan injected digitalis into the lymph sac, Behrens infused *k*-strophanthin into a vein. Evidently the frequency distribution of

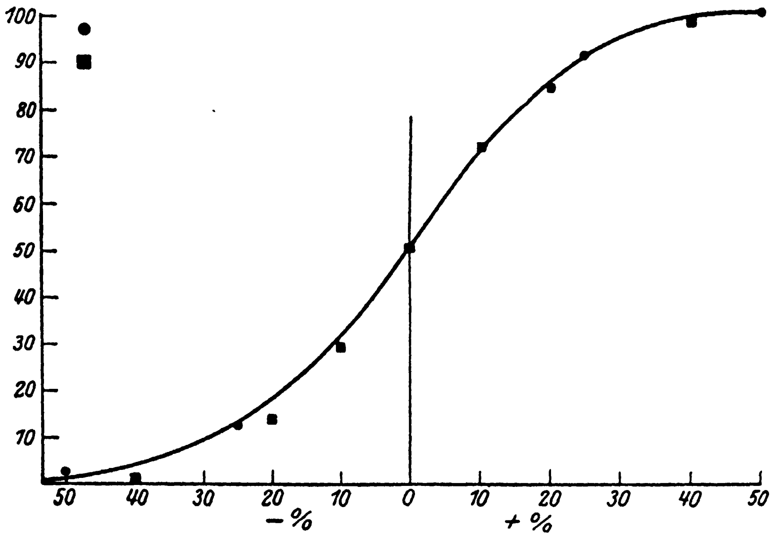


FIG. 4. Relation of dose to percentage mortality for frogs injected with ● digitalis (Trevan), ■ strophanthin (Behrens, *Arch. exp. Path. Pharmak.*, 1929, **140**, 237.)

the sensitiveness of *Rana temporaria* to digitalis is the same as to strophanthin, and Trevan is correct in speaking of the slope of the integrated frequency curve as a 'characteristic' of the species. The agreement between the results is an illustration of the beautiful regularity underlying this phenomenon of animal variation, and a testimony to the accuracy of both workers.

The construction of a dose-mortality curve. The ideal method of determining this curve is to make all the necessary observations on one day. Since this involves the use of several hundred animals it is usually impracticable, and the observations must be made on different days and then combined. Preliminary experiments are carried out to discover approximately what dose kills all frogs and what dose fails to kill, and 4 doses between these are chosen for full investigation. The effect of each

of the 4 doses should then be determined on each day observations are made; it is wrong to use all the animals which are available on one day to determine the effect of 1 dose only. The results on different days are then combined as illustrated in the following example. A sample of digitalis leaf was used to prepare the characteristic curve by injecting into frogs doses of a tincture freshly prepared from it. The results obtained on four days are given in Table IV.

TABLE IV. *The denominator gives the number of frogs injected, the numerator the number of frogs which died*

Dose (mg. leaf per gm. frog)	*1st day*	*2nd day*	*3rd day*	*4th day*
0·25	6/30	0/30	1/30	5/30
0·3	10/30	4/30	5/30	5/30
0·4	23/30	14/30	15/30	21/30
0·5	30/30	24/30	25/30	29/30

The method of combining the results may be illustrated by taking the figures for the first and second days. The figures for the first day are expressed as percentages and plotted as ordinates against the dose as abscissa; the dose causing 50 per cent. mortality is thus determined by interpolation. This dose is 0·34 mg. leaf per gm. frog. Similarly, the dose causing 50 per cent. mortality on the second day is 0·41 mg. leaf per gm. frog. The doses on each day are then expressed in terms of the dose causing 50 per cent. mortality on that day. Thus the dose of 0·25 mg. leaf per gm. frog is 0·73 of the dose causing 50 per cent. mortality on the first day. Similarly, the dose of 0·5 mg. leaf per gm. frog is 1·22 of the dose causing 50 per cent. mortality on the second day. The results for the first two days may then be expressed as in Table V.

TABLE V

First day		*Second day*	
Dose	*Percentage mortality*	*Dose*	*Percentage mortality*
0·73	20	0·61	0
0·88	33	0·73	13
1·17	76	0·97	47
1·47	100	1·22	80

The figures for the dose in Table V can then be taken as the abscissae for the corresponding percentage mortalities as ordinates. Thus the results for the two days are combined in one curve. Similarly, the results for the third and fourth days are combined with the above figures.

Quantitative Methods

The discussion of the work of Trevan and of Behrens has concerned observations in which the effect in any one animal cannot be measured; these observations may be made, not only with a drug like digitalis, but also with a serum like diphtheria antitoxin, with a hormone like the oestrus-producing hormone, or with a vitamin like the antineuritic vitamin. We may now turn to effects which are measurable, and consider one of the earliest methods described, namely, that for vitamin A due to Coward, Key, Dyer, and Morgan (1930).

Young rats are fed on a diet deficient in vitamin A until they cease to grow and begin to lose weight; each rat is then given a daily dose of the preparation to be examined, and the change in its weight is observed after a fixed period, which may be 3, 5, or 7 weeks. The irregularity in the growth response of these rats has long been known; Coward *et al.* record that one rat receiving 2·5 mg. of a given cod-liver oil gained 38 gm. in 10 weeks, while a second receiving 15 mg. lost 1 gm. in the same time. These workers showed that by taking groups of rats to determine the mean growth response, and administering each dose to a group, a relation between dose and growth response could be obtained. Their observations lay on the curve shown in Fig. 5.

In this method the quantitative response is the increase in weight of a rat, but the response may equally well be the increase in weight of an organ, such as the uterus, the prostate gland, the thyroid gland, or the pigeon's crop gland; all these have been used for different estimations. The quantitative response may also be time in hours, as, for example, in the estimation of extract of suprarenal cortex in the drake (Bülbring, 1937), or the degree of healing of a rachitic epiphysis (Dyer, 1931), or the degree of development of the endometrial glands of a rabbit's uterus (McPhail, 1934).

The construction of a dose-response curve. In constructing a curve relating dose to response, preliminary experiments must

be done to determine the doses which are to be used; let us suppose that there are 4 doses, *A*, *B*, *C*, and *D*. The best method is to take 4 animals from the same litter and to give dose *A* to one, dose *B* to the second, dose *C* to the third, and dose *D* to the fourth. Four animals of another litter are then chosen, and the doses similarly distributed among them. Further litters are taken until the effect of each dose is determined in 10–15 animals.

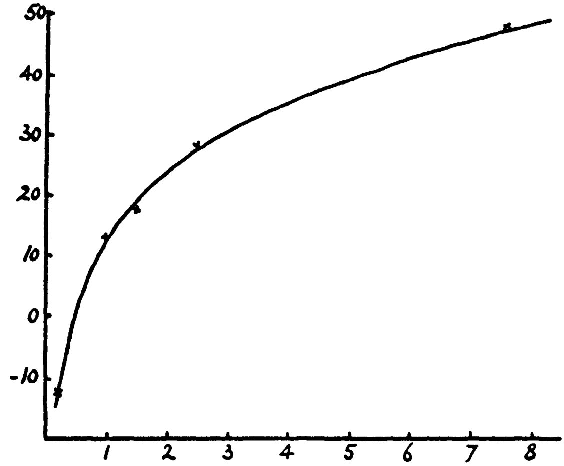


FIG. 5. Relation of the dose of cod-liver oil to the mean change in weight of groups of rats in 5 weeks. Ordinates are grammes body-weight. Abscissae are doses of cod-liver oil in mg. (Coward, Key, Dyer, and Morgan, *Biochem. J.*, 1930, **24**, 1952.)

Very large numbers of animals are rarely necessary in order to obtain a mean figure which will not vary when more animals are included. This may be illustrated by results obtained by Hume, Pickersgill, and Gaffikin, 1932. They determined the percentage of ash in the bones of rachitic rats after giving the same dose of irradiated ergosterol to each. They investigated three groups of 10 rats; although the figures for ash percentage varied as widely as from 36·7 to 52·4, the mean figures for each group agreed as well as 43·05, 43·15, and 42·32. These results are shown in Table VI.

Comparison without a curve. When the relative potency of two substances is to be determined it is now usual to work without a predetermined curve and to choose two doses of standard and two doses of unknown in the same ratio. Thus, if the two

doses of standard are 0·4 mg. and 0·3 mg., then the two doses of unknown must be in the ratio 4/3. The two doses of standard should be chosen in a range where the larger dose produces as great an increase in reponse over the response to the smaller dose

TABLE VI. *Percentage ash of the bones of rats on a rachitogenic diet receiving a daily dose of 0·01 (= 0·1 unit) of irradiated ergosterol, prophylactically for 30–35 days (Hume, Pickersgill, and Gaffikin, 1932)*

		Percentage ash in the dried fat-extracted bone		
Litter	*Sex*	*Group I*	*Group II*	*Group III*
1	♂	40·0	43·1	42·2
	♀	40·2	43·8	42·7
2	♂	51·6	41·7	39·8
	♂	42·4	36·8	42·1
	♀	38·1	41·8	36·7
3	♂	44·5	52·4	47·6
	♂	. .	. .	41·7
	♀	45·1	45·8	. .
	♀	49·6	50·6	51·1
4	♂	41·3	36·7	41·3
	♀	37·7	38·8	38·0
Average		43·05	43·15	42·32

as possible. The two doses of unknown should be chosen to produce similar effects to those produced by the two doses of standard.

Let S_1 be the smaller dose of standard
,, S_2 ,, larger ,, ,,
,, U_1 ,, smaller ,, unknown
,, U_2 ,, larger ,, ,,

then
$$\frac{S_2}{S_1} = \frac{U_2}{U_1}.$$

The mean effect of each dose is then observed in groups of animals.

Let S_1 produce a mean response f_1
,, S_2 ,, ,, ,, f_2
,, U_1 ,, ,, ,, f_3
,, U_2 ,, ,, ,, f_4.

The calculation is based on the assumption that the response is proportional to the logarithm of the dose. First we obtain from the results a measure of the effect on the response of increasing the dose; in other words, a measure of the log dose-response line. We take the mean response to the two high doses, which is $\frac{f_2+f_4}{2}$. We find the mean response to the two low doses, which is $\frac{f_1+f_3}{2}$. Then the difference between these, which is $\frac{f_2+f_4}{2}-\frac{f_1+f_3}{2}$, is the mean effect of increasing the dose by S_2/S_1. Let us call this difference c. Next we obtain a measure of the difference in the response to the standard and that to the unknown. The difference in the response to the high doses is f_2-f_4. The difference in the response to the low doses is f_1-f_3. The mean difference is $\frac{(f_2-f_4)+(f_1-f_3)}{2}$. Let us call this d. Then since the response is proportional to log dose, $\frac{c}{d}=\frac{\log S_2/S_1}{M}$, where M is the log. of the ratio of the potency of the two preparations in the doses injected. M is thus determined. An example of this calculation is given on p. 184.

Probits. When the response is one which can be measured, the relation between log dose and response is linear. When the response is the percentage of animals in which the effect is seen, the relation between the log dose and the percentage of animals in which the response occurs is not linear, but is sigmoid. The problem therefore remains how to obtain a linear relation in a percentage response.

Gaddum proposes to do this by converting a percentage response into the deviation on the normal frequency curve to which the percentage is equivalent, so that instead of the percentage we take the normal equivalent deviation. In Fig. 6 (taken from Hemmingsen, 1933) is shown a normal frequency curve and an integrated frequency curve (dotted); the latter is a percentage-response curve. The abscissae of the normal curve are stated in terms of the standard deviation; thus the mean is 0 in the abscissa, since the deviation of the mean from the mean is 0; points to the left are negative and to the right are positive. Now the ordinate 50 per cent. on the integrated curve

corresponds to the mean on the normal curve. Hence we may convert a 50 per cent. response to the value 0 which is the deviation on the normal curve to which the percentage is equivalent. Perpendiculars through the points (mean$+\sigma$) and (mean$-\sigma$) enclose about 66·6 per cent. of the area of the normal curve. Hence about 33·3 per cent. of the area is outside these

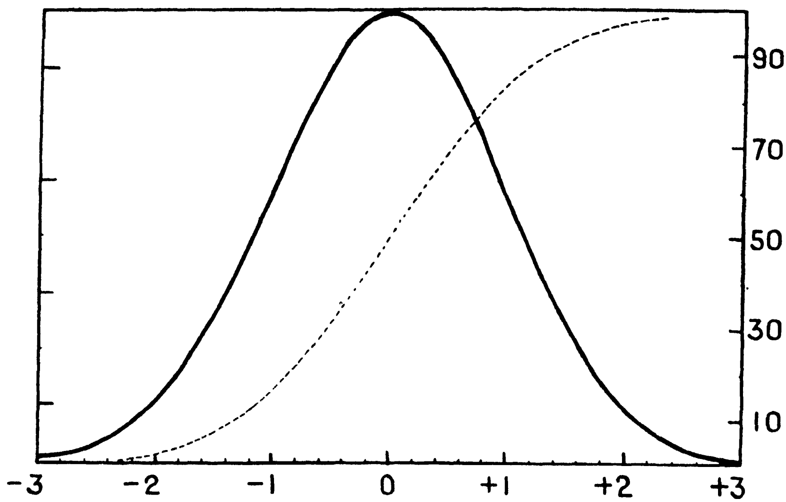


FIG. 6. To show the relation between the percentage response and the normal equivalent deviation. See text. (Hemmingsen, *Quart. J. Pharm. Pharmacol.*, 1933, **6**, 39.)

perpendiculars; since it is symmetrically disposed in two parts, the area on the left of the (mean$-\sigma$) perpendicular is about 16·6 per cent. and that on the right of the (mean$+\sigma$) perpendicular is also about 16·6 per cent. Actually a percentage response of 15·9 corresponds to a deviation of -1 on the normal curve. Similarly, a percentage response of 84·1 corresponds to a deviation of $+1$. The ordinates on the normal curve through the abscissa $-2{\cdot}0$ and $+2{\cdot}0$ enclose between them about 95·5 per cent. of the total area of the curve, and leave outside them about 4·5 per cent. Thus to the left of the ordinate through $-2{\cdot}0$ is an area corresponding to about 2·25 per cent. of the total curve. Actually, the ordinate 2·27 per cent. on the integrated curve corresponds to the abscissa $-2{\cdot}0$. Similarly, the ordinate 97·72 per cent. corresponds to the abscissa $+2{\cdot}0$. The points on the

integrated curve which correspond to points on the abscissa of the frequency curve are given in Table VII.

TABLE VII

Integrated curve	*Normal equivalent deviation*	*Probit*
2·28 per cent.	−2·0	3·0
15·9 ,,	−1·0	4·0
50·0 ,,	0	5·0
84·1 ,,	+1·0	6·0
97·72 ,,	+2·0	7·0

Bliss (1934) has suggested that to avoid the negative values, the figures for normal equivalent deviation should be added to 5·0, and then called Probits; this is a concertina word from 'Probability Unit'.

The probits for different percentages can be obtained from Fig. 7, by adding 5·0 to the normal deviation. Thus the probit for 25 per cent. is 5·0—0·68, which is 4·32. Similarly the probit for 62 per cent. is 5·0+0·3, which is 5·3.

Comparison without a curve. To compare an unknown solution with a standard preparation, the procedure is now similar to that given on p. 18. Two doses of standard are chosen likely to cause a percentage response less than 50, and a percentage response more than 50, respectively. Similarly, two doses of unknown are chosen *in the same ratio*, likely to cause similar effects. The four doses are administered each to a group of animals to determine the percentage response. This percentage response is converted to its probit, and the calculation then proceeds on the assumption that log dose is in linear relation to probit. The calculation is the same as that on p. 19, inserting probits for the mean response in a group of animals.

Discordant results. The mistake which is commonly made in estimations of this kind is that the worker deliberately excludes unusual results. If, for example, the minimal lethal dose of a sample of digitalis is determined on five cats, it often happens that four of the figures lie close together, while the fifth differs considerably. The worker is tempted to neglect the discrepant result and to take the mean of the four results which agree. Let us consider what is meant by the true value of the lethal dose

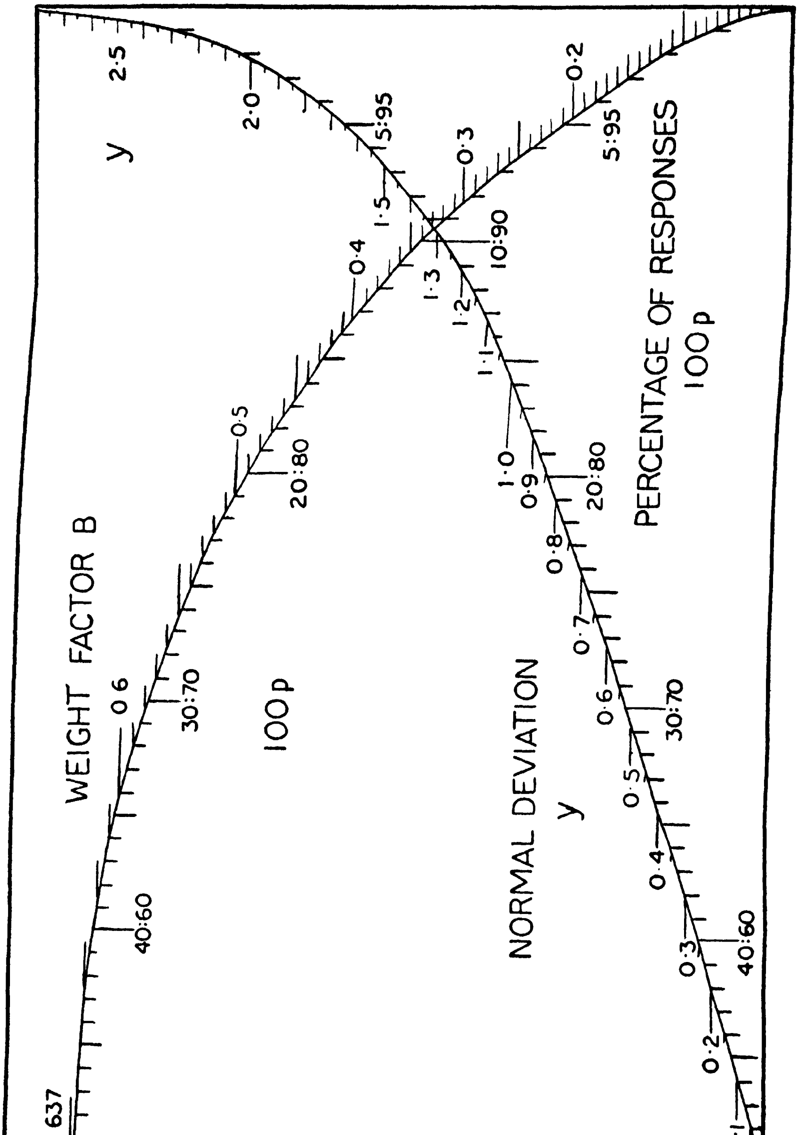


FIG. 7. Abscissae—percentage of positive effects (percentage mortality). Ordinates—the normal equivalent deviation y, and the weight factor B. When the percentage ($100p$) is greater than 50, y is positive, and when $100p$ is less than 50, y is negative. B is always positive. (Gaddum, *Med. Res. Coun.*, Spec. Rep. Ser., No. 183.)

of a sample of digitalis. The true value is the mean of a very large number of determinations, so large a number that the inclusion of other determinations does not alter the mean. If determinations are made on one animal after another, the mean of all the results obtained will steadily come nearer to the true value as the number of results increases. Consequently it can never be *assumed* that the mean of a smaller number of results is nearer to the true value than the mean of a larger number. The mean of five results, of which one differs from the other four, is likely to be nearer to the true value than the mean of the four similar results alone. That this view is correct will be generally admitted; nevertheless, in practice it is often disregarded by those who are fully aware of it: results thereby become less accurate. To prefer four results to five is merely to guess.

The pooling of material. When the effect of an agent in a group of animals is required, it often happens that much labour can be saved by pooling material from each animal in the group. Thus in estimating the effect of insulin on the blood-sugar of a group of rabbits, the samples of blood taken from the same animal at different times can be mixed together and a determination of the sugar in the mixture is then made. Similarly, in estimating a change in some constituent of urine, e.g. in the calcium content after injection of parathyroid hormone, the urine from the different animals can be mixed together, and the estimation is made on the mixture.

The cross test. Unless the number of animals in a group is very large, the mean effect of the same dose of the same agent will not be the same in two different groups, and consequently there will be an error in comparing the effect of a preparation in one group with the effect of the standard in a second group. When the experiment is one which can be repeated on each animal, the error is greatly diminished by performing what has become known as a 'cross test'. Having tested the standard on one group and the preparation on a second group on one day, the groups are changed round on a second day so that the standard is tested on the second group and the preparation on the first group. Both the standard and the preparation have then been tested on each animal in the two groups, and the mean effect of each substance is as far as possible independent of the

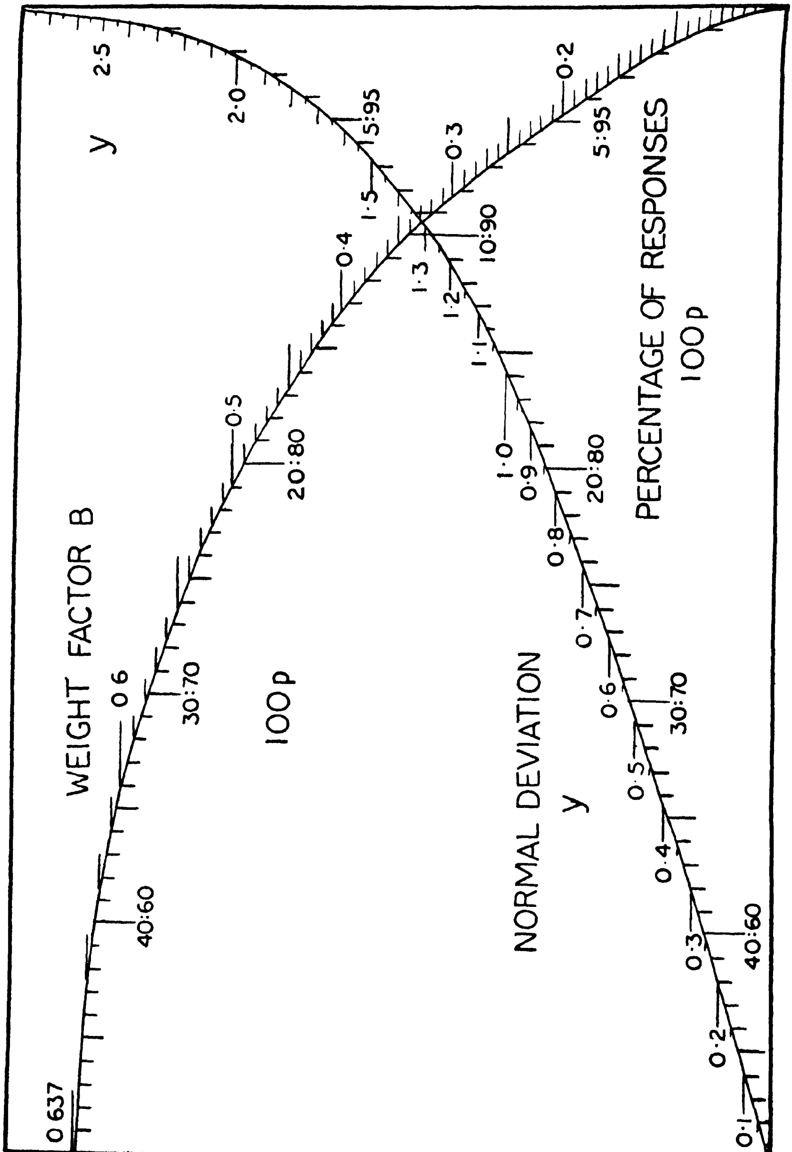


FIG. 7. Abscissae—percentage of positive effects (percentage mortality). Ordinates—the normal equivalent deviation y, and the weight factor B. When the percentage ($100p$) is greater than 50, y is positive, and when $100p$ is less than 50, y is negative. B is always positive. (Gaddum, *Med. Res. Coun.*, Spec. Rep. Ser., No. 183.)

of a sample of digitalis. The true value is the mean of a very large number of determinations, so large a number that the inclusion of other determinations does not alter the mean. If determinations are made on one animal after another, the mean of all the results obtained will steadily come nearer to the true value as the number of results increases. Consequently it can never be *assumed* that the mean of a smaller number of results is nearer to the true value than the mean of a larger number. The mean of five results, of which one differs from the other four, is likely to be nearer to the true value than the mean of the four similar results alone. That this view is correct will be generally admitted; nevertheless, in practice it is often disregarded by those who are fully aware of it: results thereby become less accurate. To prefer four results to five is merely to guess.

The pooling of material. When the effect of an agent in a group of animals is required, it often happens that much labour can be saved by pooling material from each animal in the group. Thus in estimating the effect of insulin on the blood-sugar of a group of rabbits, the samples of blood taken from the same animal at different times can be mixed together and a determination of the sugar in the mixture is then made. Similarly, in estimating a change in some constituent of urine, e.g. in the calcium content after injection of parathyroid hormone, the urine from the different animals can be mixed together, and the estimation is made on the mixture.

The cross test. Unless the number of animals in a group is very large, the mean effect of the same dose of the same agent will not be the same in two different groups, and consequently there will be an error in comparing the effect of a preparation in one group with the effect of the standard in a second group. When the experiment is one which can be repeated on each animal, the error is greatly diminished by performing what has become known as a 'cross test'. Having tested the standard on one group and the preparation on a second group on one day, the groups are changed round on a second day so that the standard is tested on the second group and the preparation on the first group. Both the standard and the preparation have then been tested on each animal in the two groups, and the mean effect of each substance is as far as possible independent of the

animal variation. This method was first recommended by Marks (1926) for the rabbit test for insulin; it has been suggested by Laqueur (1935) for the testing of oestrus-producing substances on rats or mice; it has also been adopted by Burn (1931) in the testing of pituitary (posterior lobe) extract for antidiuretic activity on rats.

Steepness of curve. No matter what the biological method, whether it be one in which the response is quantitative or in which it is a percentage, the value of the method will depend on the rate at which the effect increases in relation to the increase of the dose. This is illustrated for percentage responses when the characteristic curve for digitalis in frogs, shown in Fig. 4, is compared with the characteristic curve for the oestrus-producing hormone in rats or mice shown in Fig. 36. Whereas the ratio of LD 100 to LD 0 is 4·8 for the digitalis characteristic, it is approximately 10 for the oestrogen characteristic; it follows that the increase in effect for a given increase in dose is much greater for digitalis than for the oestrogen, and that a given number of animals will give a more accurate result for digitalis than for the oestrogen, provided that the percentage response is about the same. Similarly, when the response is quantitative, the more rapid the increase in effect for a given increase in dose, the more accurate the method.

The lesson to be drawn from this observation is that in devising a new biological method it is unwise to be satisfied with results unless the curve relating effect to dose shows a reasonably steep rise.

Simplicity of methods. Methods are good if they are accurate, rapid, and simple, and bad if they are inaccurate, slow, and need skill. That accuracy is necessary all are agreed. That speed in obtaining a result is important is not always clear to workers in academic laboratories, but it is well known to those engaged in estimations on behalf of commercial firms. Still fewer people recognize that methods are good in proportion as the technique is simple. The potency of the thyrotrophic hormone of the anterior pituitary can be estimated by the effect on the thyroid of the guinea-pig demonstrated histologically, and also by the effect on the weight of the thyroid. To estimate the effect by determinations of weight is quicker and more accurate than by histological examination; the weight method is there-

fore the better method. All biological estimations must be carried out in face of the error due to animal variation; it is, therefore, unwise to risk making the error still greater by using a more complicated technique than is necessary, which may give rise to unobserved mistakes.

The objection is sometimes raised to a method in which an organ is to be weighed, that an increase in weight is not a specific effect, whereas another method, e.g. a histological method, will show the specific change. The reply to this objection is that the quantitative estimation of a physiologically active substance should, as far as possible, be separated from qualitative identification. If identification is necessary it should be carried out first of all, but actually the source of the material is, as a rule, a sufficient guide.

REFERENCES

BEHRENS, 1929, *Arch. exp. Path. Pharmak.* **140**, 237.
BLISS, 1934, *Science*, **79**, 38 and 409.
BÜLBRING, 1937, *J. Physiol.* **89**, 64.
BURN, 1931, *Quart. J. Pharm. Pharmacol.* **4**, 517.
COWARD, KEY, DYER, and MORGAN, 1930, *Biochem. J.* **24**, 1952.
DYER, 1931, *Quart. J. Pharm. Pharmacol.* **4**, 503.
GADDUM, 1933, *Medical Research Council.* Special Report Series, No. 183.
HUME, PICKERSGILL, and GAFFIKIN, 1932, *Biochem. J.* **26**, 488.
LAQUEUR, 1935, *Klin. Woch.* **14**, 339.
DE LIND VAN WIJNGAARDEN, 1926, *Arch. exp. Path. Pharmak.* **113**, 40.
MARKS, 1926, *Publications of the League of Nations Health Organization*, C.H. 398.
MCPHAIL, 1934, *J. Physiol.* **83**, 145.
TREVAN, 1927, *Proc. Roy. Soc., B*, **101**, 483.

CHAPTER III

STATISTICAL ANALYSIS

I. Introductory Remarks

1. Statistical science. The foundation of statistical science is the common-sense handling of numerical data. Every biologist, however low he may rate his mathematical ability, is capable of understanding the logical foundations of methods of statistical inference and their application to problems of estimation, significance tests, sampling, and experimental design; that is not to say that he should necessarily himself be skilled in the use of the more complicated statistical techniques, for these may be left to the professional statistician. Too often instruction in statistics for scientists has aimed at providing them with a series of formulae into which they are expected to substitute the results of their experiments, without giving any understanding of the reasoning behind the formulae: in consequence, a formula is used because it happens to be known rather than because it is appropriate to the data. The user of statistical formulae who lacks any knowledge of the logic of his procedure is as likely to reach false conclusions as would be a chemist unable to read the labels on his bottles. Furthermore, by uncritical adherence to obsolete techniques, he will delay the extension to his science of improvements in method devised by the statisticians. It has been well said that: 'Unfortunately, too many people like to do their statistical work as they say their prayers—merely substitute in a formula found in a highly respected book written a long time ago.'

If the design of an experimental or observational programme has been well planned, the main conclusions from the numerical results can usually be derived without difficult statistical analysis. A first reading of this chapter may seem to some to contradict this assertion, but they should remember that the labour involved in a calculation is not the same thing as the difficulty: to find the product of 11973·2 and 0·09763 is more laborious than to find the product of 2·7 and 1·9, but it is no more difficult. Undoubtedly some statistical calculations are laborious, particularly for those unfamiliar with them, but that labour is

usually a small part of the work of an experiment. Indeed, the scientist well grounded in statistical principles can accomplish much by simple graphical and numerical methods, though a statistician may be able to extract additional information from the same data by more refined techniques.

2. Purpose of chapter. This chapter has a dual purpose. From a rapid reading, those who imagine themselves quite unable to make their own statistical analyses will gain a general impression of the principles and aims of statistical science in relation to biological standardization; this should prevent them from overlooking statistical difficulties entirely, and may persuade them to consult a statistician at the right stage of their investigations, during the preliminary planning, instead of only when the results have been collected. For those who are prepared to undertake a more careful study, the writer has tried to give, in non-mathematical terms, an account of elementary statistical methods sufficiently detailed for the reader to learn to apply them while remaining aware of his own limitations. Misapplication of statistics is sometimes more dangerous than non-application: those who know a little of the 'why' as well as the 'how' of statistical methods are least likely to fall victim to the danger, since they will recognize the circumstances in which they ought to seek guidance from the expert on matters outside their own experience.

The chapter is intended to be read consecutively, at least on the first occasion. The reader who is interested only in the analysis of assays based on quantal responses, for example, is warned that he will not understand Section X unless already familiar with most of Sections II–VI, as well as VIII and IX. Certain paragraphs, however, dealing with topics more difficult than the rest of the chapter, are not essential to the main themes; these, which are marked (†), may be omitted at the first reading. Paragraphs and equations have been numbered for ease of reference, and new words and phrases are italicized at their first appearance or definition. Section XIII contains a summary of the standardized algebraic notation used in the chapter.

3. Text-books. To compress into a single chapter all the statistical ideas that are important in biological standardization is impossible. Any who wish for more extensive accounts of the application of statistical science to biology should study

Fisher's well-known books (1946*a*, *b*). Mainland (1938) provides an excellent introduction for the complete novice. Snedecor (1946) presents a very full account of statistical methods with many carefully described examples of analyses, and, though he writes with particular reference to agriculture, all biologists can benefit from his book. Mather's book (1946) is written more from the point of view of the laboratory scientist, but is *perhaps slightly more difficult. Finney (1947a; especially Chapters 1–5)* gives a detailed account of the probit method for use with percentage response data. Fisher and Yates's collection of statistical tables (1948) contains a valuable introductory description of various methods of analysis, and is almost indispensable to a biological laboratory. More specialized work on the statistics of biological assays is spread over many scientific papers, references to a few of which are given at the end of this chapter.

Since the text of this chapter was completed, Emmens (1948) has published the first book that gives a comprehensive account of the statistical techniques required in biological standardization. No attempt will be made to incorporate detailed references to Emmens's book into this chapter, but the reader who wishes for further information on the design and analysis of biological assays is recommended to study it.

II. Frequency Distributions

4. Probability. An example of the variation in the sensitivity of animals to a drug has been given in Chapter II. The results obtained by de Lind van Wijngaarden showed lethal doses of digitalis in 573 cats ranging from about 42 per cent. below the mean to about 54 per cent. above the mean. The percentage of cats for which the lethal dose fell within a particular range, a quantity greatest near the mean and rapidly decreasing towards either limit of variation, was shown graphically in Fig. 1, a type of diagram known as a *histogram*. The irregularities, which are very evident in this diagram, might be expected to become relatively less important if the total number of animals were increased. Indeed, experience shows that if the number of observations contributing to such a histogram is increased very greatly, while at the same time the vertical scale is contracted so as to keep the total area constant, the diagram approaches

the form of a smooth curve or *frequency distribution*, such as that shown in Fig. 8; an area between two ordinates of this frequency distribution (e.g. the shaded area $ABVU$ in Fig. 8) represents the proportion of observations for which the measurement recorded lies between A and B. If scales are so chosen that

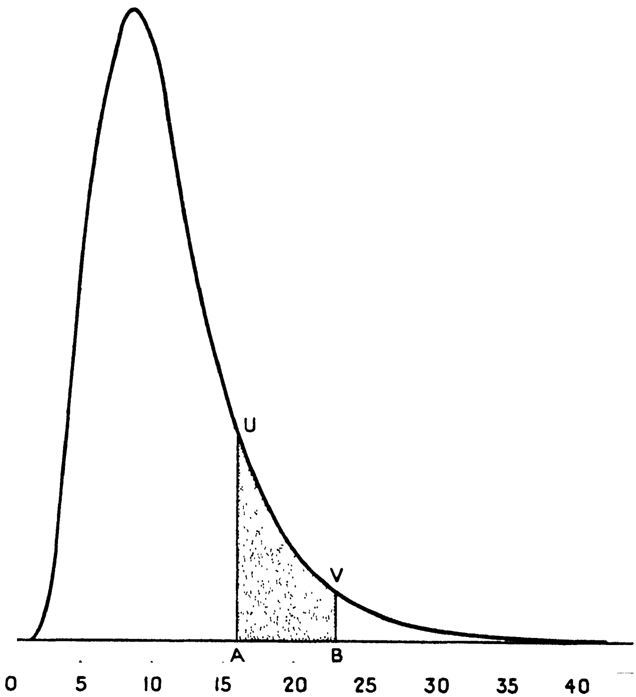


FIG. 8. A frequency distribution (compare Fig. 1). The total area between the curve and the horizontal axis is unity, and the area shaded is then the proportion of individuals for which the measurement lies between A and B.

the total area between the curve and the horizontal axis is unity, this area may be defined as the *probability*, or limiting value of the relative frequency, with which a single animal, chosen at random from the whole population, will have a value between A and B.

5. The normal distribution. Many biological measurements follow, at least approximately, what the statistician calls a *normal frequency distribution*—though departure from this should not be thought necessarily an indication of 'abnormality'

in the ordinary sense of the word! All normal distributions are symmetrical, and are the same shape, in the sense that a change in the horizontal scale, with a compensating change in the vertical scale so as to keep the total area unity, can transform any one normal curve into any other.

6. Mean and standard deviation. A normal distribution is completely specified by two quantities, the *mean*, or average

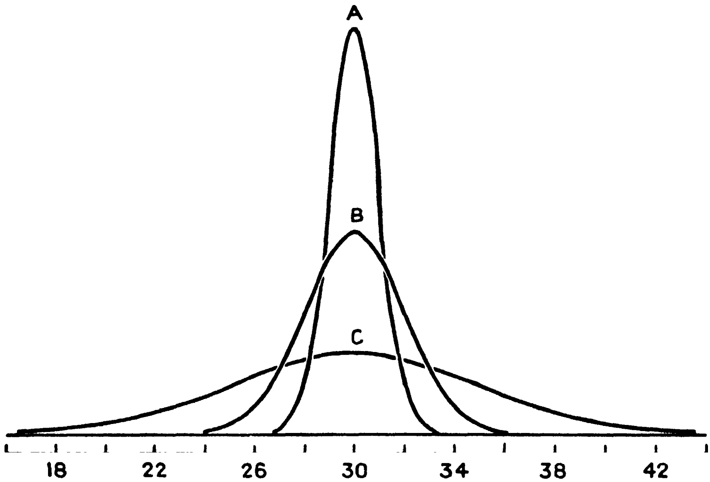


FIG. 9. Three normal distributions, all with mean 30; *A*, *B*, and *C* have standard deviations of 1, 2, and 5 respectively.

value of all the measurements so distributed, and the *standard deviation* (often abbreviated as S.D.), an expression of the spread or variability about the mean. Three normal distributions are illustrated in Fig. 9. All have the same mean, 30, and differ only in their standard deviations, which are 1, 2, and 5. The first, therefore, has only half the spread of the second and, in order to satisfy the convention that the total area shall be unity, double the height of the second: halving the horizontal scale and doubling the vertical would transform the second into the first. Because of this similarity of shape, the frequency of occurrence of individuals whose deviations from the mean lie between stated multiples of the standard deviation is the same for every normal distribution. Values have been worked out mathematically, and are available in published collections of tables (e.g. Fisher and Yates, 1948, Table I); examples are given

in Table VIII. Roughly one-third of all measurements will deviate from the mean by one standard deviation or more, and, because of the symmetry of the distribution, one-sixth will be *less* than the mean by at least one standard deviation, one-sixth will *exceed* the mean by at least this amount. In the distribution represented by curve *A* of Fig. 9, one-sixth of the measurements are less than 29 and one-sixth are more than 31, whereas in distribution *C* one-sixth are less than 25 and one-sixth more than 35. Again, roughly one-twentieth of all measurements will deviate from the mean by two standard deviations or more. Any other multiple of the standard deviation can be taken: thus, perpendiculars at 2·6 times the standard deviation on either side of the mean enclose about 99 per cent. of the curve, and the probability that an observation will lie outside these limits is about 1 in 100. The importance of results of this kind in the drawing of inferences about the mean of several observations will become clear from the paragraph 'Standard error of a mean' in Section III (§ 11).[1]

TABLE VIII. *Frequency of deviations from the mean of a normal distribution*

Deviations at least as great as this multiple of the S.D.	*Will be recorded for this proportion of all measurements*
0·67	50%, or 1 in 2
1·0	32%, or approx. 1 in 3
1·6	11%, or approx. 1 in 10
1·96	5·0%, or 1 in 20
2·0	4·6%, or approx. 1 in 20
2·6	0·93%, or approx. 1 in 100
3·3	0·097%, or approx. 1 in 1,000
3·9	0·0096%, or approx. 1 in 10,000

7. Lognormal distributions. Many classes of measurement have been found to follow a normal distribution much more closely if they are expressed in logarithmic units. The dose of a certain drug just lethal to a certain species of animal, for example, may not be normally distributed when stated as actual weight or concentration, but the logarithms of the doses may conform closely to a normal distribution. Fig. 8 is a *log-*

[1] At one time the term *probable error* was used to denote the deviation corresponding to a frequency of 1 in 2 (in a normal distribution, about 0·67 times the S.D.); the name is misleading, and the term should not be used.

normal distribution (Gaddum, 1945); its *skewness*, shown by t extended tail on the right-hand side, is typical of lethal do distributions. Fig. 1, indeed, suggests for de Lind van Wij gaarden's data a normal distribution of doses, but Bliss ar Hanson (1939) have pointed out that it includes heterogeneoι results from different samples of digitalis. The second an third diagrams in de Lind van Wijngaarden's paper, based o more restricted data, show asymmetry (not unlike that o Fig. 8) which mostly disappears on transformation to logarithms

III. Estimation and Tests of Significance

8. Estimation of a mean. The biologist who studies the digitalis tolerance of cats or the effect of vitamins on the weights of rats makes measurements on only a few of all the animals that might have been used. He examines a *sample* selected from a large *population* of animals. Obviously he cannot be sure that the mean for, say, 5 animals will coincide with the (unknown) mean of the very large population of animals that he has not measured, nor that, if he took another sample of 5 animals, these would have the same mean as the first set. Nor does he know the standard deviation in the population from which his sample is taken, though internal evidence from the sample can tell him something about this and hence about the 'repeatability' of his results.

Consider, for example, the series of numbers 13, 18, 14, 25, 14. The mean or average value of this sample of 5 is given by the familiar calculation

$$(13+18+14+25+14)/5 = 84/5 = 16{\cdot}8.$$

In symbols, if y is any one of a set of n measurements, and if Sy is used to denote the sum of all values of y, the mean may be defined as

$$\bar{y} = \frac{Sy}{n}; \qquad (1)$$

addition of a 'bar' to a symbol is a convenient method of indicating the mean value, and the symbol is now read 'y bar'.[1] If the data for this example represent measurements on 5 animals selected at random from a large population, the sample mean, 16·8, may be regarded as an *estimate* of the unknown population mean. Repetition of the process of taking samples

[1] Some writers use Σ, instead of S, to denote summation.

of 5 and calculating the mean would give different values of $\bar{y}$, some of which would be greater and some less than the true mean, but, since the latter is unknown, $\bar{y}$ gives the most trustworthy indication (at least for a normal distribution) of its value that is obtainable from one sample.

9. Spread of a sample. Now 16·8 is the mean of many other series of 5 figures, for example, of 30, 19, 23, 7, 5, or of 17, 17, 16, 18, 16, which differ from one another in the extent to which they 'spread'; the spread can be used to indicate the degree of agreement that might be found in repeated samples of 5 measurements of the same kind. For example, the first series spreads from 13 to 25, a *range* of 12, while the ranges for the second and third series are 25 and 2 respectively. The relative magnitudes of these ranges agree with the common-sense inference that a repetition of the experiment on the material from which the second series came might give a very different result, whereas on the material from which the third series came little variation in individual values would be expected and a second mean close to 16·8 would probably be found; the first series is intermediate between these extremes. A more exact expression of this property is clearly very desirable.

10. Variance and standard error. The range of a series of values makes use only of the highest and lowest members, and a better measure of variation is obtained by estimating the standard deviation. For mathematical reasons that lie beyond the scope of this discussion, the estimate is usually based on the sum of the squares of individual deviations from the mean. If y is a typical member of a sample which contains n measurements, division of the sum of squares of the deviation of the several y values from $\bar{y}$ by $(n-1)$ gives an estimate of the *variance* of the distribution:

$$s^2 = \frac{S(y-\bar{y})^2}{n-1}. \tag{2}$$

The estimated standard deviation is s, the square root of s^2. For example, the first series of figures in the last paragraph has $n = 5$, $\bar{y} = 16{\cdot}8$, and therefore

$$\begin{aligned} s^2 &= \{(-3{\cdot}8)^2+(1{\cdot}2)^2+(-2{\cdot}8)^2+(8{\cdot}2)^2+(-2{\cdot}8)^2\}\div(5-1) \\ &= 98{\cdot}80/4 \\ &= 24{\cdot}70. \end{aligned}$$

Hence the estimate of standard deviation is $s = 4{\cdot}97$; it often referred to as the *standard error* (or S.E.) of a sing determination.[1] One only of the 5 values of y deviates fro 16·8 by more than s, a reasonable agreement with the propo tion of 1 in 3 shown in Table VIII, and also a rough check o arithmetical accuracy which would reject the grosser erroı such as writing $s = 0{\cdot}497$ or $s = 49{\cdot}7$. Of course, s is only aı estimate of the true standard deviation, and, as will be seeı below, Table VIII is no longer exactly applicable.

11. **Standard error of a mean.** An important property of the normal distribution must now be introduced. If a large numbeı of samples of n are taken independently from a normal distribution, the sample means will also be normally distributed; the mean of the new distribution is the same as that of the old, the standard deviation of the new is that of the old divided by $\sqrt{n}$. Using $s_{\bar{y}}$ to denote the estimated standard deviation of the mean, this may be written

$$s_{\bar{y}}^2 = s^2/n. \tag{3}$$

For example, a large number of samples of 25 might be taken from a normal distribution which had mean 30 and standard deviation 5, the distribution shown as curve C in Fig. 9. If the mean of every sample were calculated, these would differ amongst themselves and would form a distribution: the theorem states that the new distribution would also be normal, with mean 30 but with standard deviation $5/\sqrt{25}$ or 1, and therefore, in fact, curve A in Fig. 9. Clearly the derived distribution becomes steadily narrower as the size of the sample is increased, so that any specified deviation from the population mean becomes a progressively greater multiple of the standard deviation of the sample mean, and therefore less and less likely to be exceeded. The occurrence of a single observation of magnitude 32, differing from the mean by only 0·4 times the

[1] It is usual to speak of a 'standard deviation' when s (or σ, the population value) is regarded as describing the nature of the distribution, but of a 'standard error' when the effectiveness of a sample for estimating a population mean is under examination. For the mean of a sample, the standard error concept is the one more commonly required, but for a single observation either term may be appropriate, according to circumstances. The distinction is not of great importance, except perhaps to the mathematician, and the two names may be regarded as almost synonymous.

standard deviation, would be no cause for surprise; a sample of 25 independent observations with a mean of 32, however, would constitute a deviation from the mean of twice the standard deviation, an event so rare as to occur about once in 20 trials, and a sample of 100 with a mean deviating from 30 by so much as 2 units would occur less than once in 10,000 trials. Herein lies the justification for basing conclusions on a mean of a number of observations in preference to a single observation: a mean of several observations may be expected to be nearer to the true mean of the populations studied than is a single observation. Furthermore, under very wide conditions, even though the distribution of single observations is not exactly normal, the distribution of the mean of n observations approaches more and more closely to the normal form as n increases. Inferences based on the theory of normal distributions may safely be drawn from samples of, say, 10 or more observations, even though they might be seriously misleading if applied to single observations.

12. Degrees of freedom. The estimates of variance, s^2, and of standard deviation, s, in the numerical example of § 10 are said to be based on 4 *degrees of freedom* (d.f.). The concept of degrees of freedom, which frequently recurs in statistical work, is not easy to explain in an elementary manner, but may be regarded as representing the number of independent differences, or comparisons, between the observations. If the total number of observations were 2, all the information about the spread of the distribution would be contained in one comparison, the difference between them. For 3 observations, there are effectively only 2 independent comparisons; the mean has to be calculated from the observations, and, if any 2 of the deviations $(y-\bar{y})$ are known, the third is determined as the sum of the 2 with the sign changed. In general, n observations all relating to the same variate yield a standard error with $(n-1)$ degrees of freedom.

The reader who understands a geometrical model more easily than an algebraic argument may think of the number of degrees of freedom as a number of dimensions. The values of y for a sample of 3 observations can be taken as the coordinates of a point in three-dimensional space. Knowledge of the mean of the sample is equivalent to the restriction of the point to variation on a plane (i.e. a two-dimensional region) instead of within the whole space. A sample of n observations can be

Hence the estimate of standard deviation is $s = 4{\cdot}97$; it is often referred to as the *standard error* (or S.E.) of a single determination.[1] One only of the 5 values of y deviates from 16·8 by more than s, a reasonable agreement with the proportion of 1 in 3 shown in Table VIII, and also a rough check on arithmetical accuracy which would reject the grosser errors such as writing $s = 0{\cdot}497$ or $s = 49{\cdot}7$. Of course, s is only an estimate of the true standard deviation, and, as will be seen below, Table VIII is no longer exactly applicable.

11. Standard error of a mean. An important property of the normal distribution must now be introduced. If a large number of samples of n are taken independently from a normal distribution, the sample means will also be normally distributed; the mean of the new distribution is the same as that of the old, the standard deviation of the new is that of the old divided by $\sqrt{n}$. Using $s_{\bar{y}}$ to denote the estimated standard deviation of the mean, this may be written

$$s_{\bar{y}}^2 = s^2/n. \tag{3}$$

For example, a large number of samples of 25 might be taken from a normal distribution which had mean 30 and standard deviation 5, the distribution shown as curve C in Fig. 9. If the mean of every sample were calculated, these would differ amongst themselves and would form a distribution: the theorem states that the new distribution would also be normal, with mean 30 but with standard deviation $5/\sqrt{25}$ or 1, and therefore, in fact, curve A in Fig. 9. Clearly the derived distribution becomes steadily narrower as the size of the sample is increased, so that any specified deviation from the population mean becomes a progressively greater multiple of the standard deviation of the sample mean, and therefore less and less likely to be exceeded. The occurrence of a single observation of magnitude 32, differing from the mean by only 0·4 times the

[1] It is usual to speak of a 'standard deviation' when s (or σ, the population value) is regarded as describing the nature of the distribution, but of a 'standard error' when the effectiveness of a sample for estimating a population mean is under examination. For the mean of a sample, the standard error concept is the one more commonly required, but for a single observation either term may be appropriate, according to circumstances. The distinction is not of great importance, except perhaps to the mathematician, and the two names may be regarded as almost synonymous.

standard deviation, would be no cause for surprise; a sample of 25 independent observations with a mean of 32, however, would constitute a deviation from the mean of twice the standard deviation, an event so rare as to occur about once in 20 trials, and a sample of 100 with a mean deviating from 30 by so much as 2 units would occur less than once in 10,000 trials. Herein lies the justification for basing conclusions on a mean of a number of observations in preference to a single observation: a mean of several observations may be expected to be nearer to the true mean of the populations studied than is a single observation. Furthermore, under very wide conditions, even though the distribution of single observations is not exactly normal, the distribution of the mean of n observations approaches more and more closely to the normal form as n increases. Inferences based on the theory of normal distributions may safely be drawn from samples of, say, 10 or more observations, even though they might be seriously misleading if applied to single observations.

12. Degrees of freedom. The estimates of variance, s^2, and of standard deviation, s, in the numerical example of § 10 are said to be based on 4 *degrees of freedom* (d.f.). The concept of degrees of freedom, which frequently recurs in statistical work, is not easy to explain in an elementary manner, but may be regarded as representing the number of independent differences, or comparisons, between the observations. If the total number of observations were 2, all the information about the spread of the distribution would be contained in one comparison, the difference between them. For 3 observations, there are effectively only 2 independent comparisons; the mean has to be calculated from the observations, and, if any 2 of the deviations $(y-\bar{y})$ are known, the third is determined as the sum of the 2 with the sign changed. In general, n observations all relating to the same variate yield a standard error with $(n-1)$ degrees of freedom.

The reader who understands a geometrical model more easily than an algebraic argument may think of the number of degrees of freedom as a number of dimensions. The values of y for a sample of 3 observations can be taken as the coordinates of a point in three-dimensional space. Knowledge of the mean of the sample is equivalent to the restriction of the point to variation on a plane (i.e. a two-dimensional region) instead of within the whole space. A sample of n observations can be

represented by a point in n-dimensional space, and again knowledge of the mean restricts the point to variation in an $(n-1)$-dimensional region.

13. The *t* distribution. For the numerical example, then, the mean, $\bar{y} = 16{\cdot}8$, may be regarded as arising from a normal distribution whose standard deviation is estimated, on a basis of 4 degrees of freedom, to be $s_{\bar{y}}$, where

$$s_{\bar{y}}^2 = \frac{24{\cdot}70}{5}$$

$$= 4{\cdot}94,$$

and therefore $$s_{\bar{y}} = 2{\cdot}22.$$

This result is written

$$\bar{y} = 16{\cdot}8 \pm 2{\cdot}22,$$

the '$\pm$' (read 'plus or minus') serving to introduce a standard deviation or standard error. If this $s_{\bar{y}}$ were derived from a known true standard deviation all would be well, and further inferences could be drawn by means of Table VIII. Since it is only an estimate, itself subject to errors of estimation, a modified table known as the 'Distribution of t' (Appendix II)[1] must be used. This table extends the information in Table VIII so as to apply to the problem now under discussion. For instance, Table VIII states that a deviation of more than 1·96 times the standard deviation occurs with a probability of 0·05 (or 1 in 20); the column headed 0·05 in Appendix II shows how this multiple must be altered when the standard deviation is unknown and has to be replaced by a standard error estimated with a specified number of degrees of freedom. If the standard error is based on only 2 degrees of freedom, deviations exceeding 4·30 times the standard error will occur with a probability of 0·05, but as the number of degrees of freedom is increased this multiple decreases. For 30 degrees of freedom, it is 2·04, and, since an extremely large number of degrees of freedom amounts to almost complete knowledge of the true standard deviation, it approaches a limit of 1·96 as the number of degrees of freedom increases still further.

[1] The table given is a simplified version of that presented by Fisher and Yates (1948, Table III) and in many books on statistics, but it contains sufficient detail for most purposes.

In order to protect himself from falling into a common error in more complex applications of the t distribution, either later in this chapter or elsewhere, the reader should learn now that, *without exception*, the number of degrees of freedom for t is the number on which the variance estimate, s^2, used in the calculation of t is based.

14. Fiducial limits. For the numerical example, with 4 degrees of freedom, the t-deviate at 0·05 probability is 2·78; hence the probability that $\bar{y}$ deviates from the true mean by $2{\cdot}78 s_{\bar{y}}$ or more is 0·05. Since

$$2{\cdot}78 s_{\bar{y}} = 2{\cdot}78 \times 2{\cdot}22 = 6{\cdot}2,$$

if we assert that the true mean lies between (16·8−6·2) and (16·8+6·2), or 10·6 and 23·0, we are almost sure to be right: the probability that a statement derived in this way will be wrong is only 0·05. The loose statement: 'there is a probability of 0·95 that the true mean lies between 10·6 and 23·0' should be avoided, for the true mean, though unknown, has a single definite value and not a frequency distribution. A simple conventional form of words is: 'the best estimate of the true mean from the sample of 5 observations is 16·8±2·22, and the *fiducial limits*, at a probability of 0·95, are 10·6 and 23·0.' A logical point of some subtlety is involved, and the seeker after fuller understanding should read Fisher (1946*b*, § 62). The lower limit, 10·6, is the lowest value that might be suggested for the true mean without the data giving strong evidence to the contrary, and the upper limit, 23·0, is similarly the highest value that is not contradicted by the data (see § 16). In symbols these limits are

$$\bar{y} - t s_{\bar{y}} \quad \text{and} \quad \bar{y} + t s_{\bar{y}}. \tag{4}$$

Similar calculations for the two samples of five observations given in § 9 lead to means of 16·8±4·76 and 16·8±0·37 respectively. For the first, since again there are 4 degrees of freedom, fiducial limits will be set at a distance

$$2{\cdot}78 \times 4{\cdot}76 = 13{\cdot}2$$

on either side of the mean, that is to say, at 3·6 and 30·0. So for the second, the limits are 15·8 and 17·8. The difference between these three samples in respect of the range of values for the true mean which can be regarded as acceptable show how a small

standard error implies a precise estimate of the true mean, whereas a large standard error leaves the investigator in considerable doubt about where his true mean lies.

15. Calculation of sum of squares. The calculation of the estimate s^2 from the sum of squares of deviations of individual y-values from their mean has been described. This procedure would be very tedious for a long series, as the difference between $\bar{y}$ and every y would have to be recorded before the squaring could begin. A simple piece of algebra shows that, for any set of values of y,

$$S(y-\bar{y})^2 = S(y^2)-\frac{(Sy)^2}{n}, \tag{5}$$

and use of this result simplifies the calculations very considerably;[1] the squaring is now performed on the observations themselves, so avoiding the preliminary calculation of every $(y-\bar{y})$. In the example used in § 8, application of this rule gives

$$\begin{aligned} S(y-\bar{y})^2 &= 13^2+18^2+14^2+25^2+14^2-\frac{(13+18+14+25+14)^2}{5} \\ &= 169+324+196+625+196-\frac{84^2}{5} \\ &= 1{,}510-\frac{7{,}056}{5} \\ &= 1{,}510-1{,}411{\cdot}2 \\ &= 98{\cdot}8 \end{aligned}$$

as before. Division by the degrees of freedom, 4, then gives s^2. Some may think at first that this form of calculation is more laborious, rather than less, but experience shows the contrary; these very simple figures scarcely give an adequate illustration of the advantages of the modified method, which is ideally suited for speedy and easily checked computation on a calculating machine, though even a table of squares can be used with great effect. In the absence of a calculating machine, it is sometimes helpful to use the property that $S(y-\bar{y})^2$ is unaltered by the subtraction of the same quantity from every y. With the

[1] Note that Sy^2, or $S(y^2)$, is 'the sum of the squares of values of y', and $(Sy)^2$ is 'the square of the sum of values of y'. As is illustrated by the example here, the two are quite different!

figures discussed here, the squaring can be much simplified by first subtracting 10 from every y:

$$S(y-\bar{y})^2 = 3^2+8^2+4^2+15^2+4^2-\frac{(3+8+4+15+4)^2}{5}$$
$$= 330-\frac{34^2}{5}$$
$$= 98{\cdot}8,$$

or, if care is exercised in working with negative numbers, by subtracting the whole number nearest to $\bar{y}$, 17, from every y:

$$S(y-\bar{y})^2$$
$$= (-4)^2+1^2+(-3)^2+8^2+(-3)^2-\frac{(-4+1-3+8-3)^2}{5}$$
$$= 99-\frac{(-1)^2}{5}$$
$$= 98{\cdot}8.$$

These subtractions, however, provide many opportunities for arithmetical mistakes, and a check of the calculations by subtraction of a different constant is desirable.

The phrase 'sum of squares of deviations' is often abbreviated to 'sum of squares', since there is seldom any fear of confusion.

16. Statistical significance. The mean of a sample is said to differ *significantly* from any postulated value for the population mean which lies below the lower or above the upper fiducial limit. In the first numerical example discussed above, an assertion that the five observations were a sample from a population whose true mean was 23·5 would be regarded as untenable, because, if 23·5 were the true mean, so low a sample mean from five observations would be very unlikely to occur; the method of calculation of the fiducial limit ensures that the probability of that occurrence is less than 0·05. The hypothesis that the true mean was 23·5 would be rejected by this *test of significance.* Of course, the choice of a probability of 0·05, or 5 per cent., as defining the level of significance is entirely arbitrary. Anyone who feels unprepared to regard as practically impossible an event that would occur by chance once in twenty trials can quite legitimately set himself a higher standard, say, a probability of 0·01 instead of 0·05; in so doing he reduces the risk of concluding that a difference is real when it is not, but he

must realize that he pays the price of making himself less sure to detect the existence of genuine differences between his *hypothesis relating to the mean and the evidence provided by the sample. The choice rests with the experimenter, and there is no peculiar sanctity in the 0·05 probability. 'Significant at the 5 per cent. level of probability' is a phrase now widely accepted* as denoting a reasonable assurance that the effect so described is not a chance result of the variability inherent in the population sampled; nevertheless there are situations in which either a more or a less stringent test of significance is preferable. The *British Pharmacopoeia*, for example, uses a 0·01 probability and therefore adopts fiducial limits at a probability of 0·99. Many writers like to state limits for two probabilities, 0·95 and 0·99.

Tests of significance are seldom needed for data from a single series of observations; they are more important when two series are to be compared. In biological standardization, the chief interest lies in estimation and not in the mere establishment of the existence of differences. Some understanding of tests of significance is important, especially for the purpose of examining the validity of biological assays, but, in this field of application especially, the great emphasis often placed on significance tests in expositions of statistical methods is undesirable.

17. Difference of two means. As a simple example of a test of significance for comparing two samples, the data contained in Table IX may be examined. Two tinctures of strophanthus, A and B, were tested by the cat method (Chapter XIII), each being administered to seven cats. The lethal doses, expressed in units of 0·01 c.c. undiluted tincture per kg. body-weight of cat (in order to avoid a large number of decimal places), were as shown in the first two columns of Table IX. In accordance with the conventions of later sections, the letter x will be used to represent a measure of dose, so that y may be kept as a symbol for the response of test subjects. These two columns are therefore two samples of values of x. The second pair of columns contains the calculations for forming sums of squares of deviations of each series of lethal doses from its mean; if a calculating machine were used, the individual squares would not be written down but would be summed directly to give the totals at the foot of the table. From each total is subtracted

the corresponding $(Sx)^2/n$, which for A is

$$(11{\cdot}75)^2 \div 7 = 19{\cdot}7232,$$

to leave the two sums of squares of deviations, each with 6 degrees of freedom, in the last line of the table.

TABLE IX. *Lethal doses of two tinctures of strophanthus to cats*

	Lethal doses in 0·01 c.c. tincture per kg. body-weight		*Squares of lethal doses*	
	A	*B*	*A*	*B*
	1·55	2·42	2·4025	5·8564
	1·58	1·85	2·4964	3·4225
	1·71	2·00	2·9241	4·0000
	1·44	2·27	2·0736	5·1529
	1·24	1·70	1·5376	2·8900
	1·89	1·47	3·5721	2·1609
	2·34	2·20	5·4756	4·8400
Totals	11·75	13·91	20·4819	28·3227
	$(Sx)^2/n$		19·7232	27·6412
	$S(x-\bar{x})^2$		0·7587	0·6815

A property of the normal distribution, more general and more important than that stated in § 11, is that if two normal distributions have the same standard deviations (but not necessarily the same mean), and samples of n_1 observations from the first, n_2 from the second, are taken, then the difference between the means of the two samples also has a normal distribution; the mean of this distribution is the difference between the means of the two populations sampled (as might be expected) and the standard deviation is the original standard deviation multiplied by $\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$.[1] For the present data, the means of the two samples are $\bar{x}_A$ and $\bar{x}_B$, where

$$\bar{x}_A = 11{\cdot}75/7 = 1{\cdot}68$$

and $$\bar{x}_B = 13{\cdot}91/7 = 1{\cdot}99.$$

[1] If the true standard deviations of the two distributions are σ_1, σ_2, then the S.D. of the difference in means is $\sqrt{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}$, and with this modification the theorem is true whether or not $\sigma_1 = \sigma_2$. If $\sigma_1 \neq \sigma_2$, however, theoretical difficulties arise when estimated standard deviations, s_1, s_2, have to be substituted for the unknown σ_1, σ_2. The case of $\sigma_1 = \sigma_2$ is the most important in practice.

If the experiment were repeated many times, using a series of pairs of samples of seven cats, each pair would yield a difference $(\bar{x}_B - \bar{x}_A)$; the theorem states that, if the lethal doses for each tincture are normally distributed with the same standard deviation, the values of $(\bar{x}_B - \bar{x}_A)$ will be normally distributed about the difference between the true (unknown) mean lethal doses for all cats. The difference here, 0·31, is one value from this normal distribution.

Furthermore, from the theorem, the standard deviation in this distribution of differences can be estimated if an estimate, s, of the standard deviation for individual lethal doses is available. Each sample has provided 6 degrees of freedom for such an estimate, and combination of the evidence of the two samples gives

$$s^2 = \frac{S(x_A - \bar{x}_A)^2 + S(x_B - \bar{x}_B)^2}{(n_A - 1) + (n_B - 1)} \tag{6}$$

$$= \frac{0{\cdot}7587 + 0{\cdot}6815}{6+6}$$

$$= 0{\cdot}1200;$$

n_A, n_B are the numbers of lethal doses measured for A, B respectively. Since the samples are the same size, the same results would have been obtained by calculating s^2 for each separately and then averaging the two. However, addition of the two sums of squares of deviations and division by the sum of the degrees of freedom is the general rule to be followed. When the samples are unequal in size ($n_A \neq n_B$), it rightly gives greater weight to the evidence from the larger number of degrees of freedom. According to the theorem, the standard error of d, the difference between $\bar{x}_B$ and $\bar{x}_A$, is s_d, given by

$$s_d^2 = s^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right) \tag{7}$$

$$= 0{\cdot}1200 \times (\tfrac{1}{7} + \tfrac{1}{7})$$

$$= 0{\cdot}0343,$$

and therefore $s_d = 0{\cdot}185$.

Now x was measured in units of 0·01 c.c. per kg., and therefore the estimated difference in mean lethal doses is

$$d = 0{\cdot}0031 \pm 0{\cdot}00185 \text{ c.c. per kg.}$$

The argument already used, based on the t distribution, allows inferences to be drawn about the true value of this difference. Appendix II shows the t-deviate with 0·05 probability to be 2·18 for 12 degrees of freedom (the number of degrees of freedom used for calculating s^2). Now $2{\cdot}18 \times 0{\cdot}00185 = 0{\cdot}0040$, so that the fiducial limits within which the true differences in mean lethal doses may be asserted to lie are $-0{\cdot}0009$ and 0·0071 c.c. per kg. Since this range includes zero as a plausible value for the difference in means, the data cannot be regarded as significant evidence of a real difference in potency of the two preparations. The test of significance could alternatively be made by calculating from the data a value of t which expressed d as a multiple of its standard error:

$$t = \frac{d}{s_d} \tag{8}$$

$$= \frac{0{\cdot}0031}{0{\cdot}00185}$$

$$= 1{\cdot}68.$$

The statement that t is less than the tabulated value, 2·18, is equivalent to the statement that the lower fiducial limit for d is less than zero, and it therefore shows that the two means do not differ significantly. This does not imply that the true potencies of the preparations are necessarily equal, but only that the difference is too small to be detected without more extensive data.

18. Relative potency. Often more interest attaches to the ratio of the mean lethal doses, the *relative potency* of the tinctures, than to their difference. Here R, the potency of tincture A relative to tincture B, may be estimated as

$$R = \frac{\bar{x}_B}{\bar{x}_A} \tag{9}$$

$$= 1{\cdot}18;$$

that is to say, 1 c.c. of A is estimated to have the same effect as 1·18 c.c. of B. The standard error of the estimate may be obtained from the approximate formula for $V(R)$, the variance of R:

$$V(R) = \frac{s^2}{\bar{x}_A^2}\left(\frac{1}{n_B} + \frac{R^2}{n_A}\right); \tag{10}$$

the standard error is then the square root of $V(R)$. Before using the formula, however, the quantity

$$g = \frac{s^2 t^2}{n_A \bar{x}_A^2} \tag{11}$$

should be calculated, where t is the 0·05 probability deviate with the number of degrees of freedom for s^2. If g is small, say less than 0·1, the approximation may safely be used, but if g is larger than 0·1 seriously misleading conclusions may follow its use. No proof of these statements, or of analogous statements and formulae in later sections, can be given here, but Fieller (1940), Finney (1947*b*), and Irwin (1937) may be consulted for further information. Of course, if fiducial limits are to be based on a probability other than 0·05, the corresponding t-value must be inserted in equation (11).

In the present example

$$g = \frac{0{\cdot}120 \times (2{\cdot}18)^2}{7 \times (1{\cdot}68)^2}$$

$$= 0{\cdot}029,$$

sufficiently small for safety. Therefore

$$V(R) = \frac{0{\cdot}120}{(1{\cdot}68)^2} \times \left[\frac{1}{7} + \frac{(1{\cdot}18)^2}{7}\right]$$

$$= 0{\cdot}0145,$$

whence $$R = 1{\cdot}18 \pm 0{\cdot}120.$$

Again with the t-value of 2·18 (for 12 degrees of freedom), fiducial limits for the relative potency may be calculated. Since

$$2{\cdot}18 \times 0{\cdot}120 = 0{\cdot}26,$$

these are at 0·92 (= 1·18−0·26) and 1·44 (= 1·18+0·26). Tincture A has a potency almost certainly greater than 92 per cent. but less than 144 per cent. of that of tincture B.

19. Logarithmic scale of doses. The calculations in §§ 17, 18 have been based upon the assumption that the lethal doses of individual cats conform to a normal distribution. In fact, as already mentioned in § 7, the distribution of the logarithms of lethal doses is often much closer to normality than that of the doses themselves. The measurements made in a single

experiment are usually too few to discriminate between the two possibilities, but, in view of past experience, a lognormal distribution is perhaps a better basis for the statistical analysis (Gaddum, 1945). Incidentally, this has other advantages. Standard deviations for the lethal dose distribution of preparations with widely differing potencies often show approximate proportionality to mean lethal doses; the assumption that two distributions have equal standard deviations, which has been made in § 17, is then unjustifiable. When this is so, the standard deviations of the logarithms will be practically the same for all preparations, so that the theorem on differences between two normal distributions can be applied. Also the antilogarithm of the difference in the mean logarithmic doses of two preparations will be an estimate of relative potency, for it is the ratio of equally effective doses. The determination of fiducial limits to a relative potency estimated in this way therefore does not involve the difficulties encountered in the comparison of the two tinctures of strophanthus (§ 18): the limits can be calculated for the difference in log doses, instead of directly for a ratio, and the complications that enter when g is large (essentially because the denominator of the ratio has a large standard error) do not arise.

As an illustration of the calculation in logarithmic units, the strophanthus tinctures may be compared in this way also; opportunity may be taken to include comparisons with an ouabain preparation, which was tested at the same time. The logarithms of the lethal doses in Table IX, together with the additional information on lethal doses of ouabain for nine cats, are shown in Table X. Using x now to represent the logarithmic doses, the means for the three series are

$$\bar{x}_A = 0{\cdot}217,$$

$$\bar{x}_B = 0{\cdot}292,$$

$$\bar{x}_0 = 0{\cdot}774.$$

The sum of squares of deviations from the mean may be calculated for each series just as before, the three sums having 6, 6, and 8 degrees of freedom respectively. In order to obtain the most reliable variance estimate, the information from the three groups of animals should be pooled as in equation (6);

TABLE X. *Log lethal doses of two tinctures of strophanthus and of ouabain to cats*

	Log lethal doses of strophanthus in 0·01 c.c. per kg. body-weight		*Log lethal doses of ouabain in 0·01 mg. per kg. body-weight*
	Tincture A	*Tincture B*	
	0·190	0·384	0·718
	0·199	0·267	0·996
	0·233	0·301	0·678
	0·158	0·356	0·814
	0·093	0·230	0·825
	0·276	0·167	0·760
	0·369	0·342	0·693
			0·661
			0·825
Totals	1·518	2·047	6·970
Means	0·217	0·292	0·774

addition of the sums of squares and division by the total number of degrees of freedom gives

$$s^2 = \frac{0{\cdot}0467+0{\cdot}0352+0{\cdot}0879}{6+6+8}$$

$$= 0{\cdot}00849.$$

Use of such a pooled variance estimate for comparisons between two of the three preparations may seem strange, but it has the merit of utilizing all information on variance provided by the experiment. If there were evidence that the variance for one preparation differed markedly from the others, pooling would be misleading and undesirable, but here on general grounds equality is to be expected. Only data from the one experiment should be included in this process, as a variance estimated from a different experiment might not be truly comparable. The method of analysis of variance (§ 25) applied to data such as these automatically leads to a pooled estimate, and in more complex experiments this is practically unavoidable.

Again the standard error of the difference between the two strophanthus tinctures is

$$= \sqrt{\frac{2s^2}{7}}$$

$$= \pm 0{\cdot}0493.$$

The standard error now is based on 20 degrees of freedom (6+6+8), so that the 0·05 t-value is 2·09. Since the estimated difference between the two tinctures is $\bar{x}_B - \bar{x}_A = 0{\cdot}075$, *the fiducial limits, obtained by adding and subtracting*

$$2{\cdot}09 \times 0{\cdot}0493,$$

are −0·028 and 0·178; the first of these, by the usual convention for logarithms, may alternatively be written $\bar{1}{\cdot}972$. Zero is included in this range, and therefore the analysis agrees with that previously made in showing no significant evidence of a real difference in potency of the two tinctures. Furthermore, the present analysis avoids any difficulty in assessing fiducial limits of a ratio. The estimate of relative potency is the antilogarithm of the difference between mean log doses:

$$\begin{aligned} R &= \text{antilog}(\bar{x}_B - \bar{x}_A) \qquad (12) \\ &= \text{antilog}\, 0{\cdot}075 \\ &= 1{\cdot}19, \end{aligned}$$

and the antilogarithms of the fiducial limits to this difference are the fiducial limits for the relative potency, 0·94 (= antilog $\bar{1}{\cdot}972$) and 1·51 (= antilog 0·178). Tincture A is estimated to have 119 per cent. of the potency of tincture B, and there is good reason to believe that its true potency lies between 94 per cent. and 151 per cent. of that of B. These differ only slightly from the limits obtained above in the non-logarithmic analysis, 92 per cent. and 144 per cent., but the present method is much to be preferred. It should be noted that, in general, the fiducial limits so calculated will not be equally distant on either side of R.

The potency of either tincture of strophanthus relative to that of ouabain may be estimated by similar calculations. For tincture A and ouabain,

$$\bar{x}_0 - \bar{x}_A = 0{\cdot}557 \pm 0{\cdot}0464,$$

since now $$s_d = \sqrt{\{s^2(\tfrac{1}{7}+\tfrac{1}{9})\}}.$$

Consequently, the fiducial limits to the difference in log doses are 0·460 and 0·654, a range which excludes zero, and a test of significance shows (though demonstration here was scarcely necessary) that the potency of 1 c.c. of strophanthus is definitely

different from that of 1 mg. of ouabain. From the antilogarithms of $(\bar{x}_0 - \bar{x}_A)$ and its limits,

$$R = 3{\cdot}61,$$

with limits at 2·88 and 4·51; that is to say, 1 c.c. of tincture A is estimated to be the equivalent of 3·61 mg. of ouabain, and the true potency almost certainly lies between 2·88 mg. and 4·51 mg. of ouabain.

IV. Dose-response Curves

20. Meaning of relative potency. The object of biological standardization is to compare the potencies of two preparations which, when applied as stimuli to biological material (animals, animal tissue, or micro-organisms, according to circumstances) produce responses of the same type by means of the same or similar active constituents. If equally effective doses can be determined, and if either it is known *a priori* that the preparations contain the same active principle and are therefore comparable, or it is verified experimentally that the ratio of equally effective doses is independent of the response, then this ratio, the relative potency, measures the value of the one preparation in terms of the other.

The direct method of estimating a relative potency is to assign a group of subjects to each preparation and to measure, for each subject, the dose which is just sufficient to produce a specified response; the estimate is the ratio of the mean doses for the two groups. In this method, death of the subject is frequently used as the 'response'. An example of the statistical technique required has been given in §§ 17–19; strophanthus and ouabain were administered to cats in amounts just sufficient to cause death, and, from the mean lethal doses, the potency of the strophanthus was estimated to be such that 1 c.c. was equivalent to 3·61 mg. ouabain. A relative potency of two poisons, estimated as the ratio of mean doses lethal to some species of animal, may be useful as a descriptive property of the poisons, but, unless the toxic effect is known to be due to the same active principle in each, this must not be interpreted as a general statement of their relative effectiveness in all circumstances. If there are extraneous reasons for believing that the two preparations produce their responses in the same manner—

in fact that Wood's *Hypothesis of Similarity* (1946), or Finney's *Condition II* (1947*b*), is satisfied—the general interpretation of a relative potency is justified. If not, an experimental test of the inherent comparability of the preparations must be made: the frequency distributions of individual lethal doses on the logarithmic scale should be identical for the two, except for a difference in mean values equal to the logarithm of the relative potency, and a test of assay validity might be based upon the equality of variances for the two distributions. This essential condition for the validity of an assay is discussed again in § 43.

21. **Indirect assay.** One objection to this direct assessment of relative potency is the need to ensure that a dose only just sufficient to produce the response is given. Not only will there be practical difficulties in arranging this, but, if there should happen to be some lag between the administration of a dose and the production of its effect, the estimate of the mean lethal dose may be biased. Either because of this danger, or because no test subject and response suitable for direct potency estimation can be found, an alternative and indirect form of biological assay must often be used. Indirect methods of assay require that some study be made of the relationship between the magnitude of a dose and the magnitude of a quantitative response produced by it. A dose-response curve can be used to assess the dose which would produce a specified response: consequently, if dose-response curves for both preparations are known, they can be used to give a comparison of equally effective doses. The statistical procedure will be illustrated in later sections of the chapter. As a preliminary, calculations concerned with the estimation of a *linear regression equation* must be considered, for these form an integral part of any study of dose-response relationship.

22. **Data for linear regression.** As part of a study of the response to the international standard for vitamin D, measured in rats by means of the line test, the results shown in Table XI were obtained. The table gives the line-test scores for four groups of rats, each group having received a different dose of solution containing the vitamin. The score for each rat was read independently by two observers, to the nearest $\frac{1}{2}$-unit on the conventional 0 to 6 scale, and the average of the two was regarded as the final assessment of the rat. In order to have the

convenience of working with integral values, all scores have been multiplied by 4 before entry in Table XI. If it is desired to

TABLE XI. *Line-test scores (multiplied by 4) for rats dosed with vitamin D*

(Data from Dr. E. C. Wood)

	Dose per rat (i.u.)			
	0·64	1·28	2·15	4·30
	2	4	8	14
	0	9	17	14
	2	4	6	13
	4	13	14	19
	0	3	17	17
	5	7	16	17
	3	4	8	20
	4	4	..	18
	2	10	..	17
	6	12	..	..
	1	11	..	..
Totals . .	29	81	86	149
No. of rats .	11	11	7	9
Mean scores .	2·64	7·36	12·29	16·56

quote conclusions in terms of the 0 to 6 scale, this can be restored later by a division by 4, but for assay purposes the new scale is just as good as the old. The scores or responses for any one dose are obviously very variable and show a considerable amount of 'overlap'; for example, the poorest response for a dose of 2·15 i.u. was no better than the best for a dose of 0·64 i.u., yet the best was better than several responses for a dose of 4·30 i.u. Nevertheless, if score is plotted against dose, the resulting diagram clearly indicates a tendency for larger responses to occur with the larger doses; indeed, if a logarithmic dose scale is used, and the mean score for each group of rats is plotted against the logarithm of the dose (as in Fig. 10), an approximately linear relationship appears. A linear dependence of mean response upon log dose, termed by the statistician a *linear regression* of response on dose, is frequently encountered, and the technique for estimating the straight line will now be described.

23. Regression coefficient. If x denotes the logarithm of a dose and y the corresponding response, the straight line fitted to the points of Fig. 10 should be such that, over the range of doses studied, the ordinate to the line for any x will be the mean y-value that would be obtained if a large number of rats received the dose antilog x. Standard statistical theory (Fisher,

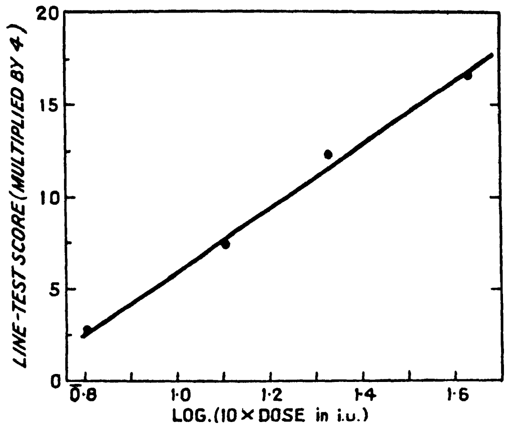


FIG. 10. Linear regression of line-test score on log dose of vitamin D. (Data in Table XI.)

1946, § 26) states that the line best representing this relationship has the equation

$$Y = \bar{y}+b(x-\bar{x}), \tag{13}$$

where $\bar{x}$, $\bar{y}$ are the mean values of x and y for the sample and b, known as the *regression coefficient*, is given by

$$b = \frac{S\{(x-\bar{x})(y-\bar{y})\}}{S(x-\bar{x})^2}, \tag{14}$$

the *sum of products of deviations* divided by the sum of squares of deviations for x. The symbol Y is used in order to avoid confusion with the particular observations, y, from which the line is estimated. Proof of these formulae requires more mathematics than would be suitable here, but some explanation can be given. The line very reasonably passes through the point $x = \bar{x}$, $Y = \bar{y}$, the mean of the observations. Now the slope of the line joining any observation to the mean (the rate of increase of y per unit increase in x in moving along the line) is

$(y-\bar{y})/(x-\bar{x})$. The regression coefficient is to be an average slope obtained from consideration of all the points, and, since points whose x is very different from $\bar{x}$ give more reliable information on the slope than those whose x is near to $\bar{x}$, the former should be given greater weight in forming the average (cf. § 87). Instead of a simple mean value of $(y-\bar{y})/(x-\bar{x})$, b is therefore taken as a weighted mean, using $(x-\bar{x})^2$ as the weight:

$$b = \frac{S[(x-\bar{x})^2 \times \{(y-\bar{y})/(x-\bar{x})\}]}{S(x-\bar{x})^2},$$

which is the expression already given. Introduction of the notation

$$S_{xx} = S(x-\bar{x})^2 \tag{15}$$

and $$S_{xy} = S\{(x-\bar{x})(y-\bar{y})\} \tag{16}$$

enables the regression coefficient to be written more concisely as

$$b = S_{xy}/S_{xx}. \tag{14}$$

Implicit in the use of this formula is an assumption that the variance of a response is constant for all values of x; this condition is unlikely to be exactly fulfilled here, but only very marked departures from it will seriously affect the conclusions.

In order to avoid negative numbers in calculation, all the logarithms of doses in Table XI will be increased by 1; this is equivalent to measuring doses in $\frac{1}{10}$ units, and the results of any analysis can easily be converted back to logarithms of the true dose at the end. The four values of x are therefore 0·806, 1·107, 1·332, and 1·633. Allowance for unequal numbers of rats in the four dose-groups leads to the mean log dose

$$\bar{x} = \frac{11 \times 0{\cdot}806 + 11 \times 1{\cdot}107 + 7 \times 1{\cdot}332 + 9 \times 1{\cdot}633}{11+11+7+9}$$

$$= \frac{45{\cdot}064}{38}$$

$$= 1{\cdot}186.$$

From the totals in Table XI

$$\bar{y} = \frac{29+81+86+149}{11+11+7+9}$$

$$= \frac{345}{38}$$

$$= 9{\cdot}079.$$

Now

$$S_{xx} = Sx^2 - \frac{(Sx)^2}{38}, \text{ by equation (5),}$$

$$= 11\times(0{\cdot}806)^2+11\times(1{\cdot}107)^2+7\times(1{\cdot}332)^2 + 9\times(1{\cdot}633)^2 - \frac{(45{\cdot}064)^2}{38}$$

$$= 57{\cdot}0457 - 53{\cdot}4412$$

$$= 3{\cdot}6045.$$

A formula analogous to equation (5) gives the sum of products of deviations:

$$S_{xy} = Sxy - \frac{(Sx)(Sy)}{38} \quad (17)$$

$$= 470{\cdot}910 - \frac{45{\cdot}064\times 345}{38}$$

$$= 61{\cdot}776,$$

the first term being the sum of the 38 products of response and corresponding dose, most readily calculated as

$$0{\cdot}806\times 29+1{\cdot}107\times 81+1{\cdot}332\times 86+1{\cdot}633\times 149.$$

Hence

$$b = \frac{61{\cdot}776}{3{\cdot}6045}$$

$$= 17{\cdot}14,$$

and the regression equation, by equation (13), is

$$Y = 9{\cdot}079+17{\cdot}14(x-1{\cdot}186).$$

If the four values of x used in the experiment are, in turn, substituted in the equation, the values of Y are 2·57, 7·72, 11·58, and 16·74; these agree well with the experimental means at the bottom of Table XI, so suggesting that the regression line (which has been drawn in Fig. 10) fits the data closely. This point is examined in § 26.

Replacement of x by $(x+1)$ converts the equation already given into one in terms of the actual doses in Table XI; after simplification, the equation becomes

$$Y = 5{\cdot}89+17{\cdot}1x,$$

the linear regression of response on log dose in i.u. The regression coefficient measures the mean increase in response per unit increase in x. It is perhaps more convenient to express this as the increase in response corresponding to a doubling of the dose. Doubling the dose is the same as increasing x by 0·301 (= log 2), so that the corresponding increase in response is $17{\cdot}14 \times 0{\cdot}301$ or 5·16. On the conventional scale for the line test, therefore, a doubling of dose is estimated to increase the score by 1·29.

24. Estimation of error. Assessment of the precision of the regression coefficient (and also of various estimates based upon the equation) requires evaluation of an error variance, s^2, which is an average value for the four doses. It may be found by forming a sum of squares of deviations of y within each dose group, adding these and dividing by the total number of degrees of freedom, (10+10+6+8); thus, for the lowest dose, the sum of squares is

$$2^2+0^2+2^2+...+6^2+1^2-\frac{29^2}{11} = 38{\cdot}55,$$

with 10 degrees of freedom. The others are formed in a similar manner to give, as in equation (6),

$$\begin{aligned} s^2 &= \frac{38{\cdot}55+140{\cdot}55+137{\cdot}43+46{\cdot}22}{34} \\ &= \frac{362{\cdot}75}{34} \\ &= 10{\cdot}67. \end{aligned}$$

The use of this estimate is illustrated in § 28.

25. Analysis of variance. An alternative to the procedure of § 24, leading to the same result but giving other valuable information, is to make an *analysis of variance*. This extremely important method of statistical analysis entails the subdivision of the sum of squares of deviations of y into components representing variation ascribable to various sources. The present data could be adequately examined without this technique; for more complex schemes of experimentation, however, some form of analysis of variance is almost essential, and a discussion of a simple example will make later uses of the method more easy to understand.

The total sum of squares of deviations of the 38 responses is

$$\begin{aligned} S_{yy} &= S(y-\bar{y})^2 \\ &= Sy^2 - \frac{(Sy)^2}{38} \\ &= 4{,}559 - 3{,}132{\cdot}24 \\ &= 1{,}426{\cdot}76 \text{ (37 degrees of freedom).} \end{aligned}$$

This is entered at the bottom of Table XII; it is a useful practice also to enter *the adjustment for the mean*, here $(Sy)^2/38$, at the top of the table, since it is generally required more than once in the analysis. A component of the sum of squares with 3 degrees of freedom can be ascribed to differences between the 4 doses. It is calculated by squaring the total response for each dose, dividing by the number of subjects at that dose, adding for the 4 doses, and subtracting the adjustment for the mean, $(Sy)^2/38$. If S_1, S_2, S_3, S_4 are used to represent summations restricted to the n_1, n_2, n_3, n_4 subjects on the 4 doses respectively, this sum of squares may be written

$$\frac{(S_1 y)^2}{n_1} + \frac{(S_2 y)^2}{n_2} + \frac{(S_3 y)^2}{n_3} + \frac{(S_4 y)^2}{n_4} - \frac{(Sy)^2}{n} \tag{18}$$

$$\begin{aligned} &= \frac{29^2}{11} + \frac{81^2}{11} + \frac{86^2}{7} + \frac{149^2}{9} - \frac{345^2}{38} \\ &= 76{\cdot}45 + 596{\cdot}45 + 1{,}056{\cdot}57 + 2{,}466{\cdot}78 - 3{,}132{\cdot}24 \\ &= 1{,}064{\cdot}01. \end{aligned}$$

Of this, a portion with 1 degree of freedom (because one *parameter* or numerical characteristic of the stock of rats, namely the regression coefficient, has been estimated) is accounted for by the regression line: this component can be proved to be

$$\frac{(S_{xy})^2}{S_{xx}} = \frac{(61{\cdot}776)^2}{3{\cdot}6045} \tag{19}$$

$$= 1{,}058{\cdot}75.$$

Algebraic proof of this is tedious. The reader may like, instead, to verify that the remainder, after the regression component has been subtracted from the sum of squares between doses, is, in fact, a sum of squares for differences between observed means and those calculated from the regression equation. Any single

response may be regarded as made up of the value given by the regression equation for that dose *plus* the deviation of the observed mean of the dose-group (shown at the foot of Table XI) from the regression value *plus* a deviation of the individual observation from the group mean; thus for the fourth response at the second dose,

$$13 = 7{\cdot}72+(7{\cdot}36-7{\cdot}72)+(13-7{\cdot}36).$$

The sum of squares for the last item has already been found (in § 24) as the sum of squares within doses: that for the second item gives the sum of squares for deviations of group means from the regression, and is

$$11\times(2{\cdot}64-2{\cdot}57)^2+11\times(7{\cdot}36-7{\cdot}72)^2+7\times(12{\cdot}29-11{\cdot}58)^2 +9\times(16{\cdot}56-16{\cdot}74)^2$$

$$= 11\times0{\cdot}0049+11\times0{\cdot}1296+7\times0{\cdot}5041+9\times0{\cdot}0324$$

$$= 5{\cdot}30.$$

The alternative calculation is

$$1{,}064{\cdot}01-1{,}058{\cdot}75 = 5{\cdot}26,$$

and the trivial discrepancy is due to the retention of only two places of decimals in the calculations. In practice, the easy subtraction would be done and the longer verification omitted.

The analysis of variance, Table XII, can now be completed by insertion of these various items. Subtraction of the sum of squares between doses from the total line leaves a sum of squares 'within doses' which is necessarily the same as that previously calculated; in the column headed 'Mean square' this is divided by its degrees of freedom to give s^2.

TABLE XII. *Analysis of variance for the data of Table XI*

Nature of variation	d.f.	Sum of squares	Mean square
Adjustment for mean		3,132·24	
Regression . . .	1	1,058·75	
Deviations from regression	2	5·26	2·63
Between doses . . .	3	1,064·01	
Within doses . . .	34	362·75	10·67
Total	37	1,426·76	

26. Test of linearity. If the analysis of variance were only a procedure for deriving a mean square within doses, it would not be needed here, for in § 24 that result has been obtained more easily by another method; even so, in many other types of experiment it is the most convenient method of achieving this end. In addition, however, the other items of the analysis give valuable information. Of particular interest here is the component measuring 'deviations from regression'. On the hypothesis that the true relationship is linear, the mean square for this must be an estimate of the variance of animals treated alike, just as is the mean square within doses. Non-linearity of regression would tend to increase the first mean square, by addition of a component of variation due to departures from the straight-line law, without affecting the second. A test of significance of the evidence against linearity may therefore be regarded as a test of whether the first mean square significantly exceeds the second. Here the issue is in no doubt, because in fact the second mean square is the larger, but the general rule is to calculate F, the ratio of the first to the second:

$$\begin{aligned} F &= s_1^2/s_2^2 \qquad (20) \\ &= 2{\cdot}63/10{\cdot}67 \\ &= 0{\cdot}2; \end{aligned}$$

this is then compared with the tabulated 5 per cent. probability values for the *distribution of the variance ratio* in Appendix III, using the entry in that table corresponding to n_1, n_2, the numbers of degrees of freedom for s_1^2, s_2^2. (In equation (20), s_2^2 is used for the quantity previously called s^2, in order to emphasize its correspondence with n_2 and to assist the application of Appendix III.)[1] For Table XII, $n_1 = 2$ and $n_2 = 34$; Appendix III does not show $n_2 = 34$, but the value will be a little less than for $n_2 = 30$. F does not approach the 0·05 probability level of about 3·3, and there is therefore no reason to doubt the hypothesis of linearity. The test of significance should always be made, for the existence of significant deviations from linearity would indicate that any further calculations based upon the

[1] The test of a variance ratio described here is identical with Fisher's z-test; in the form given (which is that now generally used) it is more conveniently and, in all but very few instances, equally satisfactorily applied.

straight line might be seriously misleading, and expert statistical advice should then be sought.

27.† Non-linearity. No attempt should be made to use a linear relationship such as that found here outside the range of doses for which it was determined. The regression of response on log dose may be effectively linear over a moderate range of doses, but will almost certainly depart from linearity for very low or very high doses. In the experiment discussed here, all responses must lie between 0 and 24, so that the regression must show some 'flattening' at extreme doses. Indeed, additional tests with a dose of 8·60 i.u. gave a mean response of 18·4 instead of the 21·9 calculated from the regression equation, and tests at a dose of 0·32 i.u. gave a mean response of 1·2 instead of *the impossible value* $-2{\cdot}6$ *from the regression equation.* This consideration does not detract in the least from the usefulness of the linear relation over the range of its validity; use of the line does not assume that it is necessarily the true dose-response law, but only that it is a satisfactory representation of that law within certain limits.

28. Variance of *b*. The variance of the regression coefficient is given by

$$V(b) = \frac{s^2}{S_{xx}} \qquad (21)$$

$$\frac{10{\cdot}67}{3{\cdot}6045}$$

$$= 2{\cdot}960,$$

so that $$b = 17{\cdot}14 \pm 1{\cdot}72.$$

Note that S_{xx} here plays the part of n in the formula for the variance of a mean. It has already been pointed out that observations for which $(x-\bar{x})$ is large give more information on the value of b than those for which $(x-\bar{x})$ is small, and estimates from single observations were weighted by $(x-\bar{x})^2$ in forming the average value; hence the divisor for the variance becomes the sum of values of $(x-\bar{x})^2$ instead of the total number of observations.

29.† Fiducial limits of *b*. The standard error of b can be used to give fiducial limits in the usual manner. The 0·05 probability level of t for 34 degrees of freedom is 2·03,

$$2{\cdot}03 \times 1{\cdot}72 = 3{\cdot}49,$$

and therefore the true regression coefficient may be asserted to lie between 13·65 and 20·63 per unit increase in x. In more easily appreciated terms, the increase in response for a doubling of the vitamin D dose is estimated to have an average value 5·16, and fiducial limits to this are 4·11, 6·21 (from multiplication by log 2).

If the reality of the statistical dependence of response on dose were in any doubt, this could be resolved either by reference to the fiducial limits of b or by a t-test. Division of b by its standard error gives

$$t = \frac{17 \cdot 14}{1 \cdot 72} = 10 \cdot 0,$$

which, when referred to Appendix II with 34 degrees of freedom, shows b to be significantly different from zero.

30. Litter-mate control. A second example, the data for which are given in Table XIII, introduces some additional important features of the analysis of variance, and may be briefly described. Bülbring and Burn (1935) investigated the effect of daily doses of oestrone on uteri of ovariectomized rats, using young females from nine litters. They adopted the device of *litter-mate control*, so as to obtain estimates of the effects of oestrone unaffected by any differences between litters in respect of their levels of response. From litters of four rats, they assigned one to each of the four levels of dose (including 'nil') to be compared; an estimated difference between two levels of dose, or an estimate of the regression coefficient of response on dose, is then affected by individual variations within litters but not by average differences in response between individuals from different litters. Unfortunately, either because three litters contained less than four females or because accidents led to the loss of several records, litter 1 gave only two results and litters 5 and 7 three each.

31. Influence of body-weight. The body-weight of a rat was thought likely to influence the weight of the uterus, independently of the dose of oestrone. The weights of uteri were therefore expressed as mg. per 100 gm. body-weight. This involves an assumption that the effect of body-weight on the uterus is one of direct proportionality; there is no strong reason for believing that such a law obtains, and a better, though more laborious, method of allowing for the influence of body-weight is that of

the analysis of covariance (Fisher, 1946*a*, § 49·1; Mather, 1946, § 34; Finney, 1948*a*). The covariance technique enables the effect of a secondary variate such as body-weight to be assessed from the internal evidence of the experiment, instead of requiring an arbitrary assumption of proportionality; the calculations are too complicated for discussion here, and could not be used for this example as only the data in Table XIII are available.

TABLE XIII. *Weights of uteri of ovariectomized rats in mg. per 100 gm. body-weight, after injection with oestrone*

(Weights for rats indicated by y_1, y_2, y_3, y_4 are not available)

Litter no.	*Daily dose for 4 days* Nil	0·2γ	0·4γ	0·8γ	*Totals*
1	42	(y_1)	124	(y_2)	166 ($+y_1+y_2$)
2	28	106	116	145	395
3	26	72	88	135	321
4	31	42	68	115	256
5	(y_3)	64	111	136	311 ($+y_3$)
6	28	70	111	133	342
7	(y_4)	56	68	85	209 ($+y_4$)
8	36	42	63	87	228
9	38	65	70	150	323
Totals	229 ($+y_3+y_4$)	517 ($+y_1$)	819	986 ($+y_2$)	2,551 ($+y_1+y_2+y_3+y_4$)

32. Missing entries. Inspection of Table XIII makes apparent that, for rats of any litter, the uterus weight increases with increasing dose of oestrone, and that there are marked differences between litters. On account of the incomplete results for three litters, simple averaging of columns of the table will not give a fair representation of the effect of dose. Litter 1 apparently contains rats with uteri above average weight, so that absence of records for this litter at doses 0·2γ and 0·8γ will tend to reduce the means for these columns relative to the others. Similarly the missing values for litters 5 and 7 will affect the comparability of the means of the first column with the others. This difficulty can be overcome by inserting numerical values for y_1, y_2, y_3, and y_4 determined in such a way that they remove any distortion of the means. Had there been only one

entry missing from the table, the formula for the appropriate numerical value (Anderson, 1946; Yates, 1933) would have been

$$y = \frac{rR+cC-G}{(r-1)(c-1)}, \tag{22}$$

where R is the total for all animals in the same litter (i.e. row of Table XIII) as the missing one, C is the total for all animals receiving the same dose (i.e. column of Table XIII) as the missing one, r and c are respectively the number of litters (rows) and the number of doses (columns), and G is the total for the $(rc-1)$ animals. Equation (22) can be seen to represent a compromise between the mean of other responses to the same dose and the mean of other responses in the same litter. It has been derived by assuming that, apart from chance variations attributable to the individuality of animals, differences between the responses of litter-mates to two doses are the same for every litter and differences between the responses of animals in two litters to the same dose are the same for every dose. If the missing entry is to be replaced by a value calculated to agree as closely as possible with this hypothesis of the purely additive character of litter and dose effects, equation (22) may be determined as the formula for the inserted value. The procedure is in fact consistent with assumptions implicit in the analysis of variance method, and involves application of the *principle of least squares*, an important general statistical technique. The formula is not a method of discovering quantities that have not been measured (though its use is often referred to as the 'calculation of a missing value'), but only a device for eliminating the bias in the comparison of means that would otherwise be caused by gaps in the table.

If more than one value is missing, an iterative form of calculation must be used. Equation (22) can be applied here to each of y_1, y_2, y_3, y_4 in turn, using guessed values for y_2, y_3, y_4 in the calculation of y_1, then using the calculated y_1 and the guessed y_3, y_4, in a revision of y_2, next the calculated y_1, y_2 and the guessed y_4 in a revision of y_3, and so on; the cycle must be repeated until further calculation does not affect the values. The end-point is independent of the initial guesses for y_2, y_3, and y_4, though good judgement of these from the appearance of the table may save much time. Verbal description of the calculations is tedious; in

practice they are easy. For example, the values of y_2, y_3, y_4 shown in the second line below might be taken as a start:

y_1	y_2	y_3	y_4
..	154	28	41
91	148	44	6
89	148	39	5
90	148	39	5

For Table XIII, $r = 9$ and $c = 4$, so that the equation (22) becomes

$$y = \frac{9R+4C-G}{24}.$$

Taking account of the guessed values for y_2, y_3, y_4 the value for y_1 should be

$$y_1 = \frac{9\times(166+154)+4\times 517-(2{,}551+154+28+41)}{24}$$
$$= 91.$$

This y_1 is used with y_3 and y_4 in a calculation of y_2

$$y_2 = \frac{9\times(166+91)+4\times 986-(2{,}551+91+28+41)}{24}$$
$$= 148.$$

The y_1 and y_2 are then used together with $y_4 = 41$ for a calculation leading to $y_3 = 44$, after which the calculated y_1, y_2, y_3 lead to $y_4 = 6$. The whole cycle of operations is repeated with these values of y_1, y_2, y_3, y_4 to give a third set, which in turn is used to give a fourth, and so on. Further calculation with $y_1 = 90$, $y_2 = 148$, $y_3 = 39$, $y_4 = 5$ will be found to give the same values, so that the process can stop there. (The reader should verify that a different set of initial values leads him to the same end-point.)

33. **Analysis of variance.** Insertion of the calculated y_1, y_2, y_3, y_4 in Table XIII leads to mean weights for each dose which are truly comparable, since any distortion resulting from unequal representation of litters has been eliminated. The adjusted totals and means are given in Table XIV. The estimate of variance used in comparing these means must be based on a residual sum of squares after elimination of a component for litter differences as well as that for differences between doses.

The calculations can be most conveniently arranged in the analysis of variance shown as Table XV. The arithmetic is similar to that which led to Table XII, but the greater symmetry, consequent upon having equal numbers of subjects for each dose, reduces the labour.

TABLE XIV. *Dose totals and means for data of Table XIII, after adjustment for missing entries*

	Daily dose				
	Nil	0·2γ	0·4γ	0·8γ	*Total*
Totals .	273	607	819	1,134	2,833
Means .	30·3	67·4	91·0	126·0	78·69

The total sum of squares is formed from the 36 entries in Table XIII, after insertion of the values for y_1, y_2, y_3, y_4. Some account must be taken of the fact that four calculated entries occur in Table XIII, and the appropriate adjustment is to reduce the total number of degrees of freedom from 35 to 31, each inserted value causing the loss of 1 degree of freedom. The sum of squares between doses (3 degrees of freedom) is found as in equation (18), by squaring the dose totals and dividing by the number of subjects; here the arithmetic is easier because there were 9 subjects for each dose, and the sum of squares is

$$\tfrac{1}{9}(273^2+607^2+819^2+1{,}134^2)-\frac{2{,}833^2}{36} = 266{,}633-222{,}941$$
$$= 43{,}692.$$

A similar calculation on litter totals gives the sum of squares between litters (8 d.f.) as

$$\tfrac{1}{4}(404^2+395^2+321^2+\ldots+323^2)-\frac{2{,}833^2}{36} = 9{,}407,$$

the divisor of the squares of litter totals being 4, the number of rats per litter. In a fully balanced experiment, with no missing entries, the sum of squares for doses would be unaffected by the magnitude of litter differences and vice versa (e.g. if all the responses observed for one litter were increased by 20, the sum of squares for doses would not be altered): the two components might be described as *independent*. Missing entries disturb this independence, but insertion of values determined by the rule

just given ensures that, as long as these form only a small proportion of the data, no serious complication is caused. The remainder after subtraction of these two components from the total,

$$57{,}912-(43{,}692+9{,}407) = 4{,}813,$$

is a sum of squares, with 20 degrees of freedom, measuring variation *'within doses and litters'*; *it is usually described more concisely as the error sum of squares.* The associated mean square,

$$s^2 = 241,$$

is the required variance per rat. The adjustment in the total number of degrees of freedom, and consequently also in the number for the error sum of squares, prevents the introduction of a bias into the estimate of variance as a result of the missing entries.

TABLE XV. *Analysis of variance for the data of Table XIII*

Adjustment for mean		222,941	
Nature of variation	*d.f.*	*Sum of squares*	*Mean square*
Regression	1	42,213	
Deviations from regression	2	1,479	740
Between doses	3	43,692	
Between litters	8	9,407	1,176
Error	20	4,813	241
Total	31	57,912	

The further subdivision of the sum of squares between doses is discussed in § 34. No particular interest attaches to the mean square for litters, except that its being significantly greater than the error mean square shows the wisdom of employing litter-mate control in planning the experiment; had the rats been assigned to the four doses completely at random, without regard to litters, a much larger variance would have been found.

Though the calculations here described lead to the correct unbiased estimate of variance per rat, the mean squares for doses and litters are not strictly appropriate to tests of significance of these sources of variation because of the introduction of the artificial y_1, y_2, y_3, y_4. The exact analysis will not be discussed here. Anderson (1946) may be consulted for information on the modifications needed; if standard errors of the

means in Table XIV were required, a good approximation would be to divide s by the square root of the number of real observations at each dose, 7, 8, 9, and 8 respectively. Of course, had no entries been missing from Table XIII, the analysis of variance calculations of this paragraph and their extensions in §§ 34, 35 would be perfectly correct (the error then having 24 d.f.), and it is from this point of view that the present account is of most importance.

34. Linearity of regression. A linear regression coefficient of weight of uterus (y) on dose (x) can be formed, by the standard method of calculation (§ 23), from

$$\begin{aligned} S_{xx} &= 9\times(0{\cdot}2^2+0{\cdot}4^2+0{\cdot}8^2)-\frac{\{9\times(0{\cdot}2+0{\cdot}4+0{\cdot}8)\}^2}{36} \\ &= 9\times0{\cdot}84-\frac{(12{\cdot}6)^2}{36} \\ &= 3{\cdot}15, \end{aligned}$$

and

$$\begin{aligned} S_{xy} &= 0{\cdot}2\times607+0{\cdot}4\times819+0{\cdot}8\times1{,}134-\frac{12{\cdot}6\times2{,}833}{36} \\ &= 364{\cdot}65. \end{aligned}$$

By equation (19), the linear regression accounts for a component of the sum of squares between doses amounting to

$$(364{\cdot}65)^2/3{\cdot}15 = 42{,}213,$$

and the remainder, with 2 degrees of freedom, provides a test of deviations from linearity. The ratio of mean squares, equation (20), is

$$F = \frac{740}{241} = 3{\cdot}1;$$

Appendix III, with $n_1 = 2$ and $n_2 = 20$, shows a significance level of 3·5, so that the evidence of non-linearity is not significant. Absence of clear evidence against linearity must not be regarded as proof of linearity, and, in view of the size of the 'deviations' mean square here, it might be wise to suspend judgement on the linearity of the regression unless and until confirmation is obtained from other experiments.

This analysis of the data in Table XIII may seem to contradict the general belief that doses should be transformed to a logarithmic scale in order that the response regression shall be linear. In fact, the data are too few to discriminate between

the alternatives. If the analysis were restricted to the animals on doses 0·2γ, 0·4γ, and 0·8γ, a linear regression on log dose would be found satisfactory, though clearly the logarithmic relationship must be wrong below 0·2γ. In view of the general experience, if the weight of uteri is to be used as a basis for oestrone assay (as exemplified in §§ 46–52), an assumption of linear regression of response or log dose seems preferable, though it may require modification for very small doses.

35. Regression coefficient. By equations (14) and (21), the regression coefficient of y on x is

$$\begin{aligned} b &= S_{xy}/S_{xx} \\ &= 364\cdot65/3\cdot15 \\ &= 115\cdot8, \end{aligned}$$

with a variance

$$\begin{aligned} V(b) &= s^2/S_{xx} \\ &= 241/3\cdot15 \\ &= 76\cdot5 \\ &= (8\cdot7)^2. \end{aligned}$$

The regression coefficient indicates an average increase in response of $115\cdot8 \pm 8\cdot7$ mg. per kg. for unit increase in dose, or $23\cdot2 \pm 1\cdot7$ mg. per kg. for 0·2γ increase in dose.

36.† Further reading. The analysis of variance is a powerful technique for the study of the effects of different sources of variation on experimental data, and those who wish to advance further in the understanding of the design and statistical analysis of biological assays should study it in detail. Bliss (1940), Bliss and Marks (1939), Bliss and Rose (1940), Fieller (1940), Finney (1948*a*), and Irwin (1937) may be consulted for examples, and the standard text-books of statistics for more formal presentation.

V. Standardization from a Linear Dose-response Regression

37. Use of standard curve. From the data of Table XI was determined an equation representing the response of rats, as measured by the line test, to doses of vitamin D. It might be thought that the equation could be used for assaying any other

preparation containing vitamin D in a very simple manner, and indeed that response curves other than straight lines could equally easily be used. Suppose that 10 rats, each of which is given a dose of 50 mg. cod-liver oil, show an average healing, on the usual line-test scale, of 2·10, or 8·40 on the scale of quarter-units used for Table XI. From the equation

$$Y = 5{\cdot}89 + 17{\cdot}1x,$$

obtained in § 23, the log dose which would give a mean response of 8·40 is

$$x = (8{\cdot}40 - 5{\cdot}89)/17{\cdot}1$$
$$= 0{\cdot}147.$$

Hence antilog 0·147, or 1·40 i.u., of standard vitamin D, as used for the tests reported in Table XI, is judged to have the same effect as 50 mg. cod-liver oil; 1 gm. cod-liver oil is estimated to contain 28·0 i.u. vitamin D.

This simple method is equivalent to measurement of the vitamin in 'rat-units', with an underlying assumption that a particular dose of vitamin will always produce the same average response. The definition of potency in terms of animal units was common in the early days of biological assay, but, as pointed out in Chapter I, it is very unsatisfactory. Unfortunately the dose-response relationship for any preparation is apt to change as a result of even small variations in experimental conditions within the one laboratory. In the line test, for example, the degree of healing is likely to be affected by the severity of rickets initially, and this will vary from one batch of rats to another. In general, it must be assumed that a dose-response curve determined on one occasion cannot be used for future standardization calculations in the way just described.[1] Still less permissible is the use of a standard curve determined in one laboratory for assays carried out in another, for the curve will almost certainly be affected by differences in the strains of animals, and differences in experimental technique will be greater than within laboratories. Consequently, if trustworthy results are to be obtained, tests of the standard pre-

[1] Only one exception has been reported: Bliss and Packard (1941) show that the curve relating X-ray dosage to percentage survival of irradiated *Drosophila melanogaster* eggs remained constant in position over a period of years, and could be used as a reference curve for standardizing dosage.

paration, S, must be run as an integral part of the assay of any test preparation, T (see §§ 43–5).

38. Use of standard regression coefficient. The response *regression is liable to shift from time to time, but, for many types of response, the shift may represent an overall change in* effective potency without any change in the sensitivity of the animals to variations in x. In other words, the mean change in response per unit increase in x (or the change in response corresponding to multiplication of the dose by a fixed factor) may remain unaltered, even though the general level of response is not constant. If experience has shown this to be a valid assumption, and a regression line for the standard preparation, S, has been determined, a test preparation, T, can be assayed relative to S by simultaneous experimentation with one dose of T and one of S. If possible, the two doses should be chosen so as to be about equivalent, as the effect on the potency estimate of any change in the regression coefficient will thereby be minimized.

Suppose that, at the same time as the test of 50 mg. cod-liver oil mentioned above, a second set of 10 rats received 5 i.u. vitamin D and showed an average healing of 2·25, or 9·00 on the scale of quarter-units. The difference in response, 0·60, corresponds to a difference of 0·60/17·1 in log dose, if the regression coefficient estimated in § 23 is assumed still to be applicable. Consequently M, the logarithm of the estimated vitamin content of 1 mg. cod-liver oil, is

$$\begin{aligned} M &= \log 5 - \log 50 - \frac{0{\cdot}60}{17{\cdot}1} \\ &= -1 - 0{\cdot}035 \\ &= \bar{2}{\cdot}965. \end{aligned}$$

The antilog of this is 0·0923: the cod-liver oil is estimated to contain 92·3 i.u. per gm., a result very different from that obtained by use of the standard curve. If y_S, y_T are the responses to log doses x_S, x_T of the standard (S) and test (T) preparations, the general formula is

$$M = x_S - x_T - \frac{y_S - y_T}{b}, \tag{23}$$

whence the relative potency is

$$R = \text{antilog}\, M. \tag{24}$$

39. Precision of *R*. On an assumption that s^2, the variance estimate found in determining the regression coefficient, remains applicable to these later responses, the variance of M *may be proved to be*

$$V(M) = \frac{s^2}{b^2}\left[\frac{1}{N_S}+\frac{1}{N_T}+\frac{(M-x_S+x_T)^2}{S_{xx}}\right], \tag{25}$$

where S_{xx} relates to the original regression calculations, and N_S, N_T are the numbers of subjects in the later tests of S and T. This formula is in some respects similar to that for the variance of a relative potency obtained directly as a ratio of mean lethal doses, equation (10). Since the expression for M involves b in the denominator, $V(M)$ can be used in the determination of fiducial limits only if b is very much larger than its standard error; the appropriate g-criterion is that

$$g = \frac{s^2t^2}{b^2S_{xx}} \tag{26}$$

should be less than 0·1. Here, since s^2 was based on 34 degrees of freedom, $t = 2{\cdot}03$ and

$$g = \frac{10{\cdot}67\times(2{\cdot}03)^2}{(17{\cdot}14)^2\times 3{\cdot}6045}$$

$$= 0{\cdot}04.$$

Hence, by equation (25),

$$V(M) = \frac{10{\cdot}67}{(17{\cdot}14)^2}\times\left[\frac{1}{10}+\frac{1}{10}+\frac{(0{\cdot}035)^2}{3{\cdot}6045}\right]$$

$$= 0{\cdot}00728,$$

and the standard error of M, the square root of $V(M)$, is $\pm 0{\cdot}085$. Multiplication of the standard error by 2·03 gives 0·173 as the interval between M and its fiducial limits, which are therefore placed at $\bar{2}{\cdot}792$ and $\bar{1}{\cdot}138$: the fiducial limits associated with the estimate of 92·3 i.u. per gm. cod-liver oil are 61·9 i.u. per gm. and 137 i.u. per gm.

40.† Combination of estimates. If two separate assays of T against S are made, the values of M should be calculated and then combined into a weighted mean. For example, if on another occasion 6 rats dosed with 2 i.u. vitamin D had a mean

line test grade of 2·17 (or 8·68 quarter-units) and 5 rats dosed with 50 mg. of the same cod-liver oil had a mean grade of 4·00 (or 16·00 quarter-units), application of equations (23) and (25) to the second series of results leads to

$$M = \log 2 - \log 50 - \frac{8{\cdot}68 - 16{\cdot}00}{17{\cdot}1}$$

$$= \bar{2}{\cdot}602 + 0{\cdot}428$$

$$= \bar{1}{\cdot}030,$$

and
$$V(M) = \frac{10{\cdot}67}{(17{\cdot}14)^2} \times \left[\frac{1}{6} + \frac{1}{5} + \frac{(0{\cdot}428)^2}{3{\cdot}6045}\right]$$

$$= 0{\cdot}01516.$$

The first estimate has the lesser variance, mainly because more subjects were tested on the first occasion, but in part because bad choice of doses on the second occasion led to a wide difference in the response for S and T and so inflated the last term in the expression for $V(M)$.

In forming a combined estimate, greater weight should naturally be given to that value of M which has the lesser standard error, for it is the more precisely determined and is expected to be the nearer to the true log relative potency. In fact, a weighted mean should be formed, using the reciprocals of the variances as weights. If M_1, M_2 are written for the separate estimates, $\bar{M}$ for the combined value,

$$\bar{M} = \frac{M_1/V(M_1) + M_2/V(M_2)}{1/V(M_1) + 1/V(M_2)} \tag{27}$$

$$= \frac{137{\cdot}4 \times \bar{2}{\cdot}965 + 66{\cdot}0 \times \bar{1}{\cdot}030}{137{\cdot}4 + 66{\cdot}0}$$

$$= -\frac{137{\cdot}4 \times 1{\cdot}035 + 66{\cdot}0 \times 0{\cdot}970}{203{\cdot}4}$$

$$= -1{\cdot}014$$

$$= \bar{2}{\cdot}986.$$

The total of the weights used in forming this weighted mean measures the whole information available on M, and, just as the individual weights were the reciprocals of the variances, so the variance of the mean value is the reciprocal of the total

weight. This may be written

$$\frac{1}{V(\bar{M})} = \frac{1}{V(M_1)} + \frac{1}{V(M_2)} \qquad (28)$$
$$= 137 \cdot 4 + 66 \cdot 0$$
$$= 203 \cdot 4,$$

whence $$V(\bar{M}) = 0 \cdot 00492.$$

The standard error of $\bar{M}$ is therefore 0·070, and, since the same s^2 was used in both assays, it has 34 degrees of freedom. Hence the fiducial limits are at $2 \cdot 03 \times 0 \cdot 070$ on either side of $\bar{M}$, or $\bar{2} \cdot 844$, $\bar{1} \cdot 128$. The antilogarithms of $\bar{M}$ and the fiducial limits then give 96·8 i.u. per gm. as the estimated vitamin content of the cod-liver oil, with fiducial limits at 69·8 i.u. per gm. and 134 i.u. per gm.

This method of combining estimates can be extended, in the obvious manner, to the combination of three or more estimates. It is applicable also to the combination of potency estimates formed as described in Sections VI, VII, and X, provided that g is not so large as to prevent the use of $V(M)$. Estimates from the self-contained assays described in these later sections, however, have each their own value of s^2, and the fiducial limits to the combined estimate cannot be found so simply. If all the values of s^2 are based on 20 degrees of freedom or more, use of a t-value of about 2·0 will be good enough in practice; if degrees of freedom are scarce, expert advice may be needed. When g is large, complexities beyond the scope of this chapter are introduced.

41. Fallacy of averages. If the same doses of each preparation had been used on the two occasions, it might seem easier to begin by averaging the responses and then to make a single calculation of M. Unless the number of subjects is the same for each group, this procedure should not be used, for it may lead the unwary into one of the most elementary statistical fallacies. The danger may be exposed by consideration of the following hypothetical figures:

	Healing grade for	
Occasion	*5 i.u. vitamin D*	*50 mg. cod-liver oil*
I	2·9 (3 rats)	2·7 (20 rats)
II	2·5 (22 rats)	2·1 (4 rats)
Totals . .	**63·7 (25 rats)**	**62·4 (24 rats)**
Mean grade .	**2·55**	**2·60**

On each occasion, the dose of cod-liver oil was less potent than that of the vitamin D, yet when the two occasions are averaged the order of potency is reversed!

42. Criticism of method. Biological assay by reference to a previously determined dose-response relationship for a standard preparation is far from satisfactory. The simplest method assumes that the standard curve remains fixed and estimates the potency of any dose of a test preparation by finding from the curve what dose of the standard would be expected to give the same response. This method has been condemned in § 37, because the assumption can seldom be justified. The second method discussed in this section involves a lesser assumption, namely, that the regression coefficient, calculated from the initial measurements of response to the standard, remains applicable in subsequent tests. Though the regression coefficient is perhaps less likely to be altered by variations in experimental conditions than is the general level of response, no theoretical reason can be adduced that it will always remain the same; a steady increase might occur, for example, if over a period the stock of laboratory animals were becoming more homogeneous as a result of selective breeding. The method of calculation illustrated also assumes that the variance of individual responses for a particular dose remains constant, though no doubt a modification could be introduced which would make use of variance estimates from the two groups of animals tested in any assay.

Types of assay which avoid these difficulties are considered in Sections VI and VII.

VI. Parallel Line Assays

43. Similar responses. Another flaw in the methods of § 37 and § 38 is that they provide no evidence on which the validity of assays may be judged. If a statement that the potency of T is 29·3 per cent. of that of S is to be any use, it must imply not merely that some particular dose of T produces the same response as would a dose of S 0·293 times the size, but that *any* dose of T could be replaced by a dose of S, calculated at this rate of equivalence, without altering the response. The process of biological standardization must in effect be a method of analysing the test preparation, so as to measure it in units of the

standard. The assay will be invalid unless the dose-response relationships for the two preparations are identical apart from a multiplier of the doses, the relative potency. For some assays, there may be strong *a priori* reasons for believing that the test and standard preparations are genuinely comparable, but even then reassurance on this point is not to be despised; for others, experimental justification of the assertion that the two produce similar responses in the test animals may be essential.

Wood (1946) has emphasized that any valid assay must be such

'that the response supposed to be produced by the known amounts of "factor X" (the substance which the investigator desires to estimate) is actually due to the factor itself and not to some other substance associated with it, e.g. an impurity; and that the response produced by the material to be analysed (assayed, or standardized) is also due solely to the presence in it of "factor X", without augmentation, diminution, or modification by any other substance also present. In other words, if we use the terms "Standard Preparation" and "Test Preparation" to denote respectively the solution of allegedly pure "factor X" and the solution prepared from the material to be analysed, we assume that the Std. Prep. contains no substance, other than factor X itself, contributing to the response we measure, and that the Test Prep. behaves for the purposes of the analysis so similarly to the Std. Prep. that it may be regarded simply as a dilution of the Std. Prep. in a completely inert diluent.'

Unless these two conditions, that of *relevance* and that of *similarity*, are satisfied, the results of any so-called assay will be meaningless and possibly misleading.

Evidence that a measured response is caused by 'factor X', and not by some alternative 'factor Y' also present in the dose, must be sought elsewhere than in the dose-response relationship. If, in fact, the responses are due to some factor present in the standard and test preparations other than that which interests the investigator, the statistical analysis necessarily produces an estimate of relative potency in respect of the other factor: no statistical procedure can demonstrate the relevance of an assay.

44. Self-contained assays. Though no positive proof of similarity can be given by the numerical results of an assay, steps can be taken to ensure that the results will betray any specified deviations from similarity. The purely logical difficulties must be surmounted in other ways; the statistical objections to the

simple assay schemes of Section V can be overcome only by making each assay a self-contained estimation process, so eliminating all need of reference to previous results (except for guidance on the most suitable doses to choose). To this end, at least two (as will be seen later, preferably at least three) doses of each preparation must be tested. The calculations required are very similar to those already described, the main difference being that a new regression coefficient is estimated for each assay from the combined evidence of the two preparations. In §§ 46–52, a simple example illustrates the procedure.

45. Tests of validity. In order that the conclusions from a biological assay shall be valid, not only must the logical conditions of relevance and similarity be satisfied, but also the assumptions underlying the form of statistical analysis adopted must be justified. For example, similarity may require (as in § 47) that two dose-response regression lines shall be parallel, while the statistical analysis adopted involves the assumptions that they are indeed straight lines, that individual responses are distributed normally about their regression lines, and that the variance of a response is independent of the dose. Some departure from strict compliance with the two last conditions can be permitted without serious effect on potency estimates and their fiducial limits, though, during the development of a new method of assay, expert statistical advice on this matter may be desirable; when a suitable statistical procedure has once been adopted, probably no interference with validity will arise from this source in the routine use of the assay.

Linearity and parallelism of regression lines are more easily disturbed by accidental variation in experimental conditions, and even small deviations can seriously upset the validity of conclusions. Though the possibility of invalidity for other reasons (either irrelevance, which is not subject to statistical control, or faulty statistical technique of the type just described) must not be forgotten, tests of assay validity are usually regarded as relating to linearity and parallelism (see § 49). In any assay such tests should be regarded as an essential feature, unless *a priori* reasons for validity are so strong as to amount to certainty. Even then, the tests provide some insurance against unforeseen flaws in the assay, such as the contamination of control diets and culture media. As will be seen from

Section XI below, in a well-planned assay little precision on the main question, the estimation of potency, need be lost because of the incorporation of tests of validity.

It must be remembered, however, that no statistical test can prove that an assay is valid, even though it may be able to demonstrate invalidity.

'While the result of any statistically invalid assay must be for that reason in error, it does not follow that the result of a statistically valid assay is not in error. If, for example, in a riboflavine assay the Test Preparation should happen to contain not only riboflavine, but also some other growth-stimulating factor, and if this factor stimulated growth proportionally to the dosage at all dosage-levels, no statistical test and no method of calculating the result could possibly detect anything suspicious in the result obtained. The combined riboflavine and other factor would be estimated as riboflavine.' (Wood, 1946.)

Before any assay is undertaken, there should be reasonable assurance of its validity, so that the tests applied may be regarded as confirmation that nothing is seriously wrong rather than as proof that everything is right.

46. Data for 4-point assay. In continuation of their work on the assay of oestrone, from which the data of Table XIII were taken, Bülbring and Burn report the results shown in Table XVI. Two doses of the standard oestrone were used, $0{\cdot}2\gamma$ and $0{\cdot}4\gamma$ daily, together with two doses of the solution of the preparation to be assayed, 0·0075 c.c. and 0·015 c.c. daily. (See § 31 for a discussion of the influence of body-weight on response.) An assay of this pattern, involving two doses of each preparation, is generally known as a *4-point design*. Bliss and Marks (1939) and Schild (1942) have given good accounts of the statistical analysis appropriate to assays of this design; Schild's is a particularly clear elementary account. Both papers, however, use notation different from that adopted here, and, in consequence, what are essentially the same methods and tests look rather different from the recommendations in §§ 47–52.

The doses of the test preparation were chosen, on the evidence of preliminary trials, as about equivalent to those of the standard. Rats from seven litters were employed in the assay, and an attempt was made to balance litter differences by assigning one animal from each litter to each of the four doses, but again the rats available were insufficient for a perfectly balanced arrangement. If the five missing entries in the table were

ignored, comparisons between doses would be biased. Equation (22), which here becomes

$$y = \frac{7C+4R-G}{18},$$

has therefore been used to calculate 'missing values'. Once again, four of the values may be guessed and the formula used to find the fifth. An iterative series of calculations, using the formula to calculate each of the five in turn, eventually produces *the values shown in brackets in Table XVI, and no further changes take place if additional calculations are based* on these. For instance, further revision of the value for the missing rat from litter 7 would give

$$y = \frac{7\times317+4\times(495+111)-(1{,}921+87+50+82+111)}{18}$$
$$= 133,$$

the value already attained. At all subsequent stages of the calculations for the assay, these five entries are included, and only approximate allowances have been made for their effect on the subsequent analysis (cf. § 33). The chief purpose of this example is to illustrate the standard calculations for a 4-point assay, and the user of the method who encounters data from which some responses are lacking would be well advised to consult a statistician.

TABLE XVI. *Weights of uteri of ovariectomized rats, in mg. per 100 gm. body-weight, after injection with oestrone*

	Daily dose				
	Oestrone		*Unknown solution*		
Litter no.	*0·2γ*	*0·4γ*	*0·0075 c.c.*	*0·015 c.c.*	*Totals*
1	54	152	61	92	359
2	49	71	74	63	257
3	51	112	51	(87)	214 (301)
4	(50)	58	60	102	220 (270)
5	81	102	(82)	120	303 (385)
6	63	(111)	83	105	251 (362)
7	126	(133)	83	108	317 (450)
Totals	424 (474)	495 (739)	412 (494)	590 (677)	1,921 (2,384)
Means	(67·7)	(105·6)	(70·6)	(96·7)	

Figures in brackets in the body of the table have been calculated. Totals without brackets exclude, totals in brackets include, the calculated values.

47. Response diagram. Though estimation of the potency of the test preparation can be made entirely by arithmetical analysis of the data, it is always advisable to draw a diagram. In Fig. 11, the mean responses from Table XVI are plotted against the log dose; separate scales are used for the two preparations, in order to bring both into a diagram of convenient

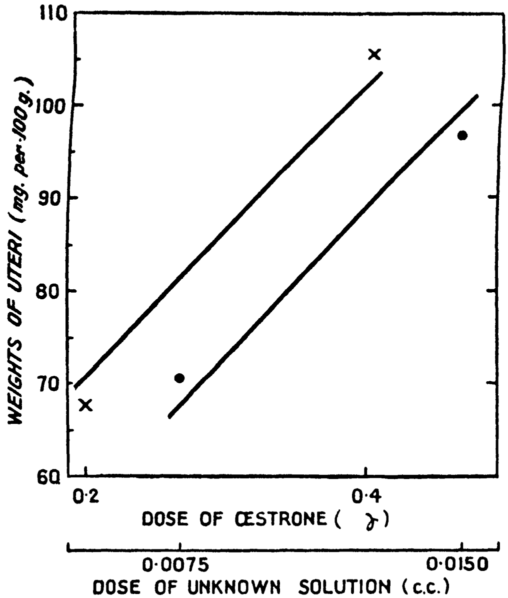


FIG. 11. Assay of oestrone from uterine weights of ovariectomized rats. (Data in Table XVI.)
× Standard preparation. ● Test preparation.
The dose scale is logarithmic, but is graduated in true doses.

size. If the test preparation produces its effect entirely as a result of its oestrone content, a doubling of the test dose must increase the log dose of oestrone by exactly the same amount as would a doubling of the dose of the standard, in fact by log 2. If a linear regression adequately represents the dependence of response on log dose of oestrone, the true regression coefficients for the two preparations must be equal (since each measures the increase in response per unit increase in log dose) and the true dose-response lines must be parallel. The results of an experiment will naturally show some departure from perfect parallelism, but this should be attributable only to chance

variations between animals. Any indication that the deviation from parallelism was greater than could reasonably be attributed to chance would amount to a contradiction of the condition of similarity (see § 49). A test of significance of non-parallelism should therefore precede any estimation of relative potency, as a confirmation that the assay is valid.

As shown in § 51, the estimate of potency is formed from the horizontal distance between the parallel lines. Consequently a graphical estimation procedure could be based upon drawing a pair of parallel lines, judged by eye to fit the four points as well as possible, and measuring the distance between them; the method will be obvious from later paragraphs, but it has the disadvantage of giving neither a test of validity nor a standard error for the estimate.

48. Analysis of variance. The first stage in the analysis proper is to complete the analysis of variance, Table XVII.

TABLE XVII. *Analysis of variance for the data of Table XVI*

Adjustment for mean . . .		202,981	
Nature of variation	*d.f.*	*Sum of squares*	*Mean square*
Preparations . . .	1	63	
Regression . . .	1	7,168	
Parallelism . . .	1	240	240
Between doses . .	3	7,471	
Between litters . .	6	7,069	
Error	13	7,165	551
Total	22	21,705	

The calculations are in the main similar to those for Table XV, but the subdivision of the sum of squares of deviations between doses is different. For example, the total sum of squares of deviations is

$$54^2+49^2+51^2+\ldots+105^2+108^2-\frac{2{,}384^2}{28}$$

$$= 224{,}686-202{,}981$$

$$= 21{,}705,$$ with 22 degrees of freedom since 5 calculated values were inserted in Table XVI.

By equation (18), the sum of squares between doses is

$$\frac{474^2+739^2+494^2+677^2}{7}-\frac{2,384^2}{28}$$

$$= 210,452-202,981$$

$$= 7,471,$$

and a like calculation gives the sum of squares between litters.

The 3 degrees of freedom between doses, and the corresponding sum of squares of deviations, can be subdivided into three independent components; of these, one depends only upon the average difference in response for the two preparations, the second depends only upon the regression coefficient of response on log dose, and the third depends only upon the difference in slope between regressions for the preparations taken separately. These can be calculated particularly easily because of the symmetrical arrangement of the experiment. The sum of squares for the difference between preparations can be found by the expression in equation (18), which might here be written

$$\frac{G_S^2}{N_S}+\frac{G_T^2}{N_T}-\frac{G^2}{N}; \tag{29}$$

G_S, G_T, G now represent the totals of responses for the standard and test preparations and for both together ($G = G_S+G_T$), N_S, N_T, N are the numbers of responses in these groups ($N = N_S+N_T$), the five calculated 'missing values' being included as if they were genuine. In numerical terms, (29) becomes

$$\frac{(677+494)^2+(739+474)^2}{14}-202,981 = 63.$$

By simple algebra it may be shown that when $N_S = N_T = \frac{1}{2}N$, a situation very common in biological assays and one that is recommended in Section XI as a principle of good design, (29) may be written

$$\frac{(G_S-G_T)^2}{N}. \tag{30}$$

This expression is more suited to rapid calculation. In the present assay, $N_S = N_T = 14$, and the required sum of squares is

$$\frac{(677+494-739-474)^2}{28} = 63.$$

For each preparation, the upper dose is twice the lower. On a logarithmic scale, the interval between them is log 2; for the lower dose of either preparation, $(x-\bar{x})$ will therefore be $-\frac{1}{2}\log 2$ and for the upper dose $+\frac{1}{2}\log 2$. For convenience in computation, these may be replaced by -1 for the lower dose, $+1$ for the upper. In order to compensate for this alteration in the scale of x, all quantities measured on the new scale (in *particular, M, its standard error, and its fiducial limits) must eventually be multiplied by $\frac{1}{2}$ log 2, or 0·1505.*[1] *On the modified* logarithmic scale, $(x-\bar{x})^2 = 1$ for every response, and therefore application of equations (15), (16), and (17), for the two preparations gives:

Standard	*Test*
$S_{xx} = 14$	$S_{xx} = 14$
$S_{xy} = 739\times(+1)+474\times(-1)$	$S_{xy} = 677\times(+1)+494\times(-1)$
$= 739-474$	$= 677-494.$

If the two regression lines are to be parallel, an average regression coefficient must be used for each; it is obtained by pooling the two values of S_{xx} and the two values of S_{xy} before application of equation (14):

$$b = \frac{677-494+739-474}{14+14}$$

$$= 16\cdot 00.$$

The regression accounts for a further degree of freedom between doses, with a contribution to the sum of squares calculated by the usual $(S_{xy})^2/S_{xx}$ formula, as in equation (19),

$$\frac{(677-494+739-474)^2}{28} = 7{,}168.$$

The close analogy between the formula for the two components now calculated suggests the form required for the third. It

[1] This may be regarded as using logarithms to base $\sqrt{2}$, instead of base 10, in order to make $(x-\bar{x})$ take the simple values -1 and $+1$. Results are later converted to base 10, so that tables of common logarithms may be used, by means of the relationship

$$\log_{10} z = \log_h z \times \log_{10} h, \qquad (31)$$

where h is any base of logarithms and z any number. Here

$$h = \sqrt{2} \text{ and } \log_{10} h = 0\cdot 1505.$$

is obtained by taking the difference between values of S_{xy} for the two preparations, instead of their sum, to give

$$\frac{\{(677-494)-(739-474)\}^2}{28} = \frac{(677-494-739+474)^2}{28}$$
$$= 240.$$

As will be seen in § 49, this component corresponds to the deviation from parallelism or the difference in regression coefficients for the two preparations, but, for conciseness, is usually named parallelism. Note that the numerators for the three components involve the three possible ways of forming a difference between two pairs of the dose totals in Table XVI. The three components add, as algebraic proof shows they must, to the sum of squares (3 d.f.) between doses.

The error mean square in Table XVII provides an estimate of variance per rat,

$$s^2 = 551,$$

to be used later in assessing the precision of R.

49. Validity test. The first two of the components into which the sum of squares between doses is analysed are of no particular interest so far as tests of significance are concerned. Their mean squares can indeed be compared with the error mean square with the aid of Appendix III, to show that the mean responses to the two preparations do not differ significantly and that the regression of response on log dose is significant: the one is valueless information until it is related to the actual doses used, and the other calls for no special comment since the assay was planned in the expectation that there would be a regression.

The third component has been formed from the differences in the two values of S_{xy}, and its magnitude depends upon the extent to which separate regressions for the two preparations deviate from parallelism. Validity of the assay has been shown to require that the true regression lines should be parallel, and a test of significance of this component is a test of assay validity. Here the variance ratio is less than unity, and there is no serious evidence of inequality of the regression coefficients; the assay may proceed, using the average b.

This argument is based upon an assumption that the true dose-response regression is linear. In fact, linearity is likely to obtain at best in a restricted range of doses: providing that the

assay is valid, the response *curves* will be parallel, but they need not be everywhere linear. If in a 4-point assay one or more of the doses used falls outside the range of linearity, an appearance of non-parallelism may be produced when the data are analysed as though the regressions were linear. The effect of invalidity of the assumptions underlying the statistical processes is thus indistinguishable from the effect of invalidity of the assay. In a 6-point assay (§ 54) the two types of invalidity can be distinguished. The importance of this for assay design is discussed in Section XI; here it is sufficient to note that when the 'parallelism' component in the 4-point analysis of variance is not significant there is no reason for alarm, but when it is significant the cause may be invalidity of statistical analysis and it is not necessarily a sign of complete invalidity of the assay (cf. § 45).

50. Regression lines. From Table XVI, the mean responses for the two preparations are

$$\bar{y}_S = (474+739)/14 = 86{\cdot}6$$

and

$$\bar{y}_T = (494+677)/14 = 83{\cdot}6.$$

With the value of b from § 48, the two regression lines to be drawn in Fig. 11 may be estimated as having the equations

$$Y_S = 86{\cdot}6+16{\cdot}00(x-\bar{x}_S),$$

$$Y_T = 83{\cdot}6+16{\cdot}00(x-\bar{x}_T);$$

these lines are shown in Fig. 11.

51. Relative potency. Once again the estimate of relative potency is the ratio of equally effective doses, or the antilogarithm of the difference between log doses that produce equal responses. The difference between equally effective x-values, M, is obtained from the two regression equations by giving Y the same value in each and evaluating the corresponding x for each. The general formula is the same as equation (23), except that mean doses and mean responses are inserted. It is

$$M = \bar{x}_S-\bar{x}_T-\frac{\bar{y}_S-\bar{y}_T}{b}, \tag{32}$$

an expression which represents the horizontal distance between the two regression lines. Here

$$M = \bar{x}_S-\bar{x}_T-\frac{3{\cdot}0}{16{\cdot}00}$$

$$= \bar{x}_S-\bar{x}_T-0{\cdot}188.$$

The logarithmic dose scale used in the analysis was described in § 48; as shown by equation (31), M must be multiplied by 0·1505 in order to give $\log R$ expressed in common logarithms (i.e. to base 10). Since the ratio of the upper doses of the two preparations is the same as that of the lower doses, $(\bar{x}_S - \bar{x}_T)$ is equal to the difference in log doses (on the modified scale used) between either pair of corresponding levels, and therefore

$$0{\cdot}1505(\bar{x}_S - \bar{x}_T) = \log\frac{0{\cdot}2}{0{\cdot}0075}.$$

Hence

$$\log R = \log\frac{0{\cdot}2}{0{\cdot}0075} - 0{\cdot}188 \times 0{\cdot}1505$$

$$= \log\frac{0{\cdot}2}{0{\cdot}0075} - 0{\cdot}028,$$

the antilogarithm of which gives

$$R = \frac{0{\cdot}2}{0{\cdot}0075} \times \text{antilog}\,\bar{1}{\cdot}972$$

$$= 25{\cdot}0.$$

Thus 1 c.c. of the unknown solution is estimated to contain 25·0γ oestrone.

52. Fiducial limits of R. The log relative potency, M, contains the regression coefficient as a denominator, and, unless the standard error of b is so small as to make zero far beyond the fiducial limits for b, complications will arise in determining fiducial limits of M. The approximate variance of M is given by equation (25), with $\bar{x}$ replacing x for each preparation

$$V(M) = \frac{s^2}{b^2}\left[\frac{1}{N_S} + \frac{1}{N_T} + \frac{(M - \bar{x}_S + \bar{x}_T)^2}{S_{xx}}\right], \qquad (33)$$

but this can be used only when $V(b)$ is small. By equation (21), taking the value of s^2 from § 48,

$$V(b) = \frac{551}{28}$$

$$= 19{\cdot}68,$$

so that b is less than four times its standard error. In general, $V(b)$ need not be calculated, as a g-criterion should be applied

in order to test whether or not equation (33) may be used. The general formula may be written just as in equation (26):

$$g = \frac{t^2 V(b)}{b^2} \tag{34}$$

$$= \frac{s^2 t^2}{b^2 S_{xx}}.$$

A value of g less than 0·1 indicates that fiducial limits may be calculated with the aid of equation (33), a value greater than unity would show that b was not significantly different from zero and the assay therefore useless. On insertion of the 0·05 level of t for 13 degrees of freedom (the number of d.f. for s^2), and other numerical values from this assay, equation (34) gives

$$g = \frac{551 \times (2{\cdot}16)^2}{(16{\cdot}00)^2 \times 28}$$

$$= 0{\cdot}359,$$

very much greater than the limit of 0·1.

This shows that the uncertainty introduced into the denominator of M by the relatively large standard error of the regression coefficient makes fiducial limits for M calculated from the formula for $V(M)$ completely unreliable. In these circumstances, a much more complicated formula for the fiducial limits is unavoidable. The lower limit for M is (Finney, 1947*b*):

$$(\bar{x}_S - \bar{x}_T) + \frac{M - \bar{x}_S + \bar{x}_T}{1-g}$$

$$- \frac{st}{b(1-g)} \sqrt{\left[(1-g)\left(\frac{1}{N_S} + \frac{1}{N_T}\right) + \frac{(M - \bar{x}_S + \bar{x}_T)^2}{S_{xx}}\right]} \tag{35}$$

$$= (\bar{x}_S - \bar{x}_T) - \frac{0{\cdot}188}{0{\cdot}641}$$

$$- \frac{23{\cdot}5 \times 2{\cdot}16}{16{\cdot}00 \times 0{\cdot}641} \sqrt{\left[0{\cdot}641 \times \left(\frac{1}{11} + \frac{1}{12}\right) + \frac{(0{\cdot}188)^2}{28}\right]}$$

$$= (\bar{x}_S - \bar{x}_T) - 0{\cdot}293 - 1{\cdot}664$$

$$= \bar{x}_S - \bar{x}_T - 1{\cdot}957.$$

Note that, in (35), N_S and N_T are taken as the number of genuine responses recorded for the two preparations, in order to make approximate allowance for the five missing values.

Hence the lower fiducial limit for the relative potency, obtained as was R in § 51, is

$$\frac{0{\cdot}2}{0{\cdot}0075}\,\text{antilog}(-1{\cdot}957\times 0{\cdot}1505) = \frac{0{\cdot}2}{0{\cdot}0075}\,\text{antilog}\,\bar{1}{\cdot}705$$
$$= 13{\cdot}5.$$

The corresponding upper limit is obtained by a formula similar to (35), in which the term involving the square root is added instead of subtracted; the result is a limit of 42·9 for R.[1] The conclusion is that the potency of 1 c.c. of the unknown solution almost certainly lies between 13·5γ and 42·9γ oestrone. The best estimate that can be formed from the data in Table XVI is 25·0γ oestrone per c.c. (§ 51), but the variability in response of different subjects is so great that this estimate might easily be wrong by almost a factor of 2 in either direction.

These limits are calculated from the internal evidence of one assay, but they and others like them are generally assumed to indicate the measure of agreement that might be expected between the results of repeated assays. The assumption is justified providing that the conditions stated in § 43 and others implicit in the method of statistical analysis adopted are completely fulfilled. Except for the effects of sampling variation, potency estimates should then be independent even of the assay technique: an assay of an insulin preparation by the rabbit blood-sugar test in one laboratory should agree with an assay of the same preparation by the mouse-convulsion test in another laboratory, for both would be estimates of the true insulin content of the test preparation. There is perhaps a tendency to assume too readily that all the conditions are satisfied, and that the variation between the results of independent assays is no greater than the internal error of one. Experimental verifications of this are few. One satisfactory instance is the series of assays of twenty-one insulin samples, using the blood-sugar method, reported by Young and Romans (1948), though unfortunately the several assays of each sample were always performed on the same day or within a few days of one another.

53.† Short-cut methods. Those who do not have to use

[1] If g were negligibly small, these formulae would be identical with those for limits based on $V(M)$, as may be seen from substitution of $g = 0$.

methods of biological standardization very frequently will find the scheme of calculations for the 4-point assay outlined above to be the most suitable, since the meaning of each stage should be clear to them. Modifications, which involve various short-cuts, may be desirable if 4-point assays are used frequently as a laboratory routine. Examples of calculations performed in this way have been given by Bliss (1946), Knudsen and Randall (1945), and Knudsen, Blackwell Smith, Vos, and McClosky (1946): they are undoubtedly quicker, but they introduce the danger of blind application of formulae mentioned in § 1.

54. **Data for 6-point assay.** As will be seen from Section XI of this chapter, it is often desirable that an assay should involve tests of more than two doses of each preparation. Bliss and Marks (1939) have discussed the statistical analysis of several more extensive symmetrical designs, and Finney (1948*a*) gives numerical details for a *6-point design*, that is to say, an assay using three doses of each preparation. Another simple example is the assay of testosterone propionate reported by Pugsley (1946), the data from which are given in Table XVIII. The technique of the assay was that of Gallagher and Koch (1935; see Chapter X below). Three doses of a standard preparation and three of a test preparation were included in the assay, each being used on five capons; the response measured is the growth of comb, conventionally taken as (length+height), of injected capons.

TABLE XVIII. *Comb growth (length+height) in mm. of capons injected with testosterone propionate*

	Dose of standard preparation			*Dose of test preparation*		
	20γ	40γ	80γ	20γ	40γ	80γ
	6	12	19	6	12	16
	6	11	14	6	11	18
	5	12	14	6	12	19
	6	10	15	7	12	16
	7	7	14	4	10	15
Totals	30	52	76	29	57	84
Means	6·0	10·4	15·2	5·8	11·4	16·8

Experience of the method and inspection of Table XVIII agree in suggesting that the regression of response on log dose

is linear. In Fig. 12 the six mean responses have been plotted against dose on a logarithmic scale; the scale for the test preparation has been displaced to the right so as to keep the two sets of points distinct. The two lines drawn in Fig. 12 are those resulting from the calculations described below, but lines almost

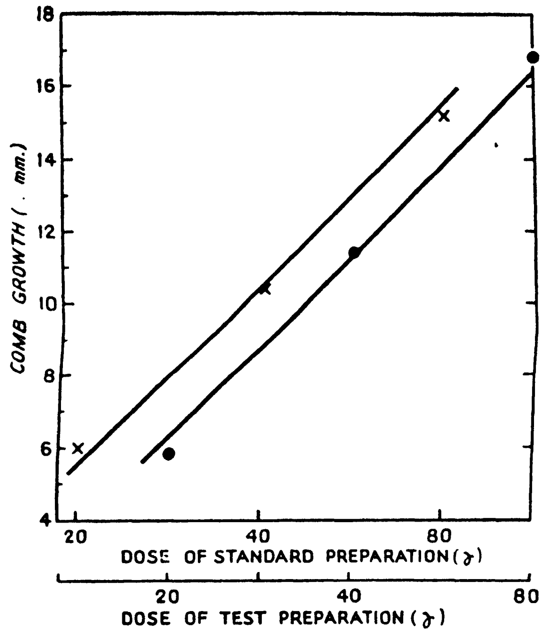


FIG. 12. Assay of testosterone propionate from comb growth of capons. (Data in Table XVIII.)

× Standard preparation. ● Test preparation.

The dose scale is logarithmic, but is graduated in true doses.

equally good might be drawn by eye, without calculation. In such a drawing the horizontal distance between the lines could be measured and used to give an estimate of R. For example, in Fig. 12 the horizontal distance may be measured as about 0·16 times the interval between doses of 20γ and 40γ (after allowing for the horizontal displacement of the scale for the test preparation). Hence, approximately,

$$M = 0{\cdot}16 \log 2 = 0{\cdot}048$$

and the antilogarithm, 1·12, is the potency of the test preparation

relative to the standard. Arithmetical analysis is needed, however, if the precision of this estimate is to be assessed.

55. Analysis of variance. The first calculations for the analysis of variance (Table XIX), leading to the formation of an error variance, s^2, follow the standard pattern. The assay has no constraint corresponding to the litter classification in the oestrone assay (§ 46), and the subdivision of the total sum of squares of deviations into 'Doses' and 'Error' is made as in § 25, with the simplification in equation (18) consequent upon having equal numbers of responses at each dose. The complete analysis of the sum of squares for doses into the five components used in the validity tests (§ 57) is more complicated, and its discussion merits a special paragraph.

TABLE XIX. *Analysis of variance of the data of Table XVIII*

Adjustment for mean		3,586·13	
Nature of variation	*d.f.*	*Sum of squares*	*Mean square*
Preparations	1	4·80	4·80
Regression	1	510·05	510·05
Parallelism	1	4·05	4·05
Curvature	1	0·02	0·02
Difference of curvatures	1	0·15	0·15
Between doses	5	519·07	
Error	24	56·80	2·367
Total	29	575·87	

56.† Subdivision of 'Doses' sum of squares. Bliss and Marks (1939) have shown the utility of sets of *polynomial coefficients* for expediting the subdivision of the sum of squares of deviations between doses, in the analysis of data of this kind; their paper should be consulted for a full account of the procedure, which is a logical extension of that used already in a 4-point assay (§ 48). The scheme of calculation for the 6-point assay is summarized in Table XX.

The meaning of Table XX is that the values given in the columns S_1, S_2, S_3, are to be used as multipliers of the total responses for the three doses of the standard preparation and the values in the columns T_1, T_2, T_3 are to multiply the total responses for the three doses of the test preparation: these total responses, copied from Table XVIII, appear in the last line of

TABLE XX. *Subdivision of sum of squares between doses in Table XIX*

(Data of Table XVIII)

Component	*Polynomial coefficients*						*Divisor*	*Sum of products*
	S_1	S_2	S_3	T_1	T_2	T_3		
Preparations .	−1	−1	−1	1	1	1	30	12
Regression . .	−1	0	1	−1	0	1	20	101
Parallelism . .	1	0	−1	−1	0	1	20	9
Curvature . .	1	−2	1	1	−2	1	60	1
Difference of curvatures .	−1	2	−1	1	−2	1	60	−3
Dose totals .	30	52	76	29	57	84		

Table XX. The sums of these five sets of products are shown in the last column of Table XX. For example, the line of Table XX labelled 'Preparations' gives

$$-30-52-76+29+57+84 = 12.$$

This is simply the overall difference in response between the two preparations, and, by formula (30), it makes a contribution to the sum of squares

$$12^2 \div 30 = 4{\cdot}80;$$

the number 30, which is N in (30), is shown in the 'Divisor' column of Table XX. So for the 'Regression' component

$$-30+76-29+84 = 101,$$

the difference between responses for the two high doses and responses for the two low doses, which is clearly in some way a measure of the effect of increase in dose on the increase in response. In fact, as will be seen from § 58, on a suitable dose scale this quantity is the sum of the S_{xy} for the two preparations, and so may be used to form the regression coefficient. The divisor for each component in Table XX may be proved to be the sum of the squares of the six coefficients multiplied by the number of subjects at each dose: for 'Regression'

$$(1^2+0^2+1^2+1^2+0^2+1^2)\times 5 = 20,$$

and therefore the sum of squares is

$$101^2 \div 20 = 510{\cdot}05.$$

The third component, 'Parallelism', or more exactly deviations from parallelism, can similarly be seen to be the difference between two values of S_{xy}, and the corresponding sum of squares, formed as in § 48 except for the modification in the rule for the divisor, is

$$9^2 \div 20 = 4{\cdot}05.$$

The fourth and fifth components cannot be so easily explained. If the regression for either preparation were perfectly linear, the total response for the high and low doses together would be equal to twice the response for the middle dose. The 'Curvature' component represents the deviation of these totals from equality, for the two preparations, and measures the average non-linearity shown by the data. The 'Difference of curvatures' is similarly formed by taking the differences between preparations. For both, the divisor is found by the same rule:

$$(1^2+2^2+1^2+1^2+2^2+1^2) \times 5 = 60.$$

The sums of squares for the five components have been entered in Table XIX, where it is apparent that they add to the total sum of squares for 'Doses'. The polynomial coefficients have, in fact, led to a subdivision of that sum of squares into five mutually independent components.[1] Some may be helped to an understanding of these calculations if they realize that the sum of squares for each component is computed just as though the polynomial coefficients for that component were the values of some x-variate to be used in a regression. For example, if with every response to a high or low dose of either preparation in Table XVIII is associated an x-value of $+1$, and with every

[1] Necessary and sufficient conditions for a set of coefficients to do this are: (*a*) that each sum of products in Table XX represents a *contrast* between doses, which means that the sum of coefficients in any one line shall be zero (e.g. for Curvatures, $1-2+1+1-2+1 = 0$), and (*b*) that any pair of contrasts shall be *orthogonal*, which means that the sums of products of corresponding coefficients shall be zero (e.g. for Parallelism and Curvature,

$$1\times 1 + 0\times(-2) - 1\times 1 - 1\times 1 + 0\times(-2) + 1\times 1 = 0).$$

Many different sets of five components could be found subject to the restriction that the coefficients for every one and for every pair in the same set satisfy these conditions; with the aid of divisors formed according to the rule already given, every set would give a subdivision of the 'Doses' sum of squares. The set in Table XX is chosen because its members are particularly relevant to the questions that will be asked of the data in an assay of this design.

response to a middle dose is associated an x-value of -2, since the sum of x for the 30 responses is zero,

$$S_{xx} = Sx^2 = 60$$

and $$S_{xy} = Sxy = 1$$

are easily verified. These are respectively the 'Divisor' and the 'Sum of Products' for Curvature in Table XX, and the usual $(S_{xy})^2/S_{xx}$ formula for regression gives the sums of squares found for this component. The orthogonality condition ensures that the effects of these five x-variates are mutually independent. As shown in § 58, the second of them is the true x-variate for log dose, but formally the other components give the same pattern of calculations.

The analysis of the 4-point assay (§ 48) may be regarded as a simpler instance of the use of polynomial coefficients, for which only three components have to be considered. The sets of coefficients were then: for preparations, -1, -1, 1, 1; for regression, -1, 1, -1, 1; for parallelism 1, -1, -1, 1. Bliss and Marks (1939) have shown how the method can be extended to 8-point and other assays; the theory of the construction of coefficients is too advanced for discussion here. When there is a cross-classification of the data according to litters or some other factor, as in the oestrone assay of § 46, the sum of squares for doses may still be subdivided by the same rules. If the design of the assay is asymmetrical, however, on account of unequal numbers of subjects at each dose, unequal numbers of doses of each preparation, or unequal spacing of log doses, the method needs modification and a statistician should be consulted.

57.† Validity tests. A significantly large mean square for parallelism would be a demonstration that the hypothesis of two parallel lines was untenable, suggesting that responses to the two preparations were not similar and the assay therefore invalid (Wood, 1946; Finney, 1947*b*). The curvature component is so chosen as to measure the average amount of deviations from linear regression in the two preparations. Its significance would indicate that a linear regression of response on log dose was an inadequate representation of the data, and calculations based on such an assumption might then be very misleading. The component described as 'difference of curvatures' provides a test of whether the amount of deviation from

linearity is different in the two preparations. Significance of one or more of these three components should be regarded as a warning that either the assay is invalid or a modified form of analysis, not dependent upon a linear regression of response on log dose, is needed; study of a diagram such as that of Fig. 12 may help greatly at this stage, but usually expert statistical advice should be sought.

58. Relative potency. *Table XIX shows none of these complications*, and the remainder of the analysis may be completed on a pattern very similar to that of the oestrone assay. On a logarithmic scale, values of $(x-\bar{x})$ for the three doses of either preparation are $-\log 2$, 0, $\log 2$. For arithmetical convenience, these may be replaced by -1, 0, 1, a procedure similar to that in § 48 and equivalent to using logarithms to base 2, providing that quantities measured on this scale are eventually multiplied by log 2 (or 0·301), in order to convert them to common logarithms by the rule of equation (31). These values of $(x-\bar{x})$ are identical with the polynomial coefficients for the regression component in Table XX. From that table, or by summation for the 30 subjects,

$$S_{xx} = 20,$$

and $$S_{xy} = 101.$$

By equation (14), $b = 5{\cdot}05.$

Consequently, by insertion of the mean responses for S and T, easily calculated from Table XVIII, the two regression lines are found to have the equations

$$Y_S = 10{\cdot}53+5{\cdot}05(x-\bar{x}_S),$$

$$Y_T = 11{\cdot}33+5{\cdot}05(x-\bar{x}_T).$$

The magnitudes of the three doses were the same for each preparation, so that

$$\bar{x}_S-\bar{x}_T = 0.$$

Hence the formula for the difference between equally effective log doses, equation (32), reduces to

$$M = \frac{11{\cdot}33-10{\cdot}53}{5{\cdot}05}$$

$$= 0{\cdot}158,$$

as compared with the value 0·16 obtained in § 54 by measurement on Fig. 12. For this assay, equation (34) gives

$$g = \frac{2{\cdot}367 \times (2{\cdot}06)^2}{(5{\cdot}05)^2 \times 20}$$
$$= 0{\cdot}02,$$

by use of $s^2 = 2{\cdot}367$ from the error line of Table XIX. Since g is so small, equation (33) may be used for the variance of M:

$$V(M) = \frac{2{\cdot}367}{(5{\cdot}05)^2} \times \left[\frac{1}{15} + \frac{1}{15} + \frac{(0{\cdot}158)^2}{20}\right]$$
$$= 0{\cdot}01248$$
$$= (0{\cdot}112)^2.$$

From Appendix II, the 0·05 level of t with 24 degrees of freedom (the number of d.f. for s^2, taken from Table XIX) is 2·06; fiducial limits to M are therefore $2{\cdot}06 \times 0{\cdot}112$ below and above M, at $-0{\cdot}073$ and 0·389 respectively. The logarithmic dose scale used involved division of the actual log doses by log 2, and therefore M and its fiducial limits must be multiplied by 0·301 before antilogarithms are taken. The result is

$$R = \text{antilog}\, 0{\cdot}0476 = 1{\cdot}12$$

with limits at antilog $\bar{1}{\cdot}9780$ and antilog 0·117, or 0·95 and 1·31 respectively. Hence 1γ of the test preparation is estimated to have the potency of 1·12γ of the standard, with 5 per cent. fiducial limits at 0·95γ and 1·31γ.

59.† Unsymmetrical assays. The two assays whose statistical analyses have been discussed were planned with a high degree of symmetry: they had equal numbers of subjects at each of two or three doses of the standard and test preparations, and the doses were equally spaced on a logarithmic scale. Such an arrangement may not always be possible, and detailed statistical analysis is then much more laborious. However, a less complete analysis, based on the same principles, will often suffice to bring out the main features of the results. The first stage of the analysis of variance, subdivision of the total sum of squares into components between and within doses, presents no difficulty (cf. § 25). From the 'between doses' component can be extracted a single degree of freedom representing the difference

between preparations, for which the sum of squares is calculated by formula (29), since (30) is applicable only when $N_S = N_T$. The average regression coefficient, and its sum of squares, can be calculated as in either of the assays described in this section, by forming values of S_{xx} and S_{xy} for each preparation and pooling them. Polynomial coefficients will not be applicable, and calculation of S_{xx}, S_{xy} will require the method of § 25 for the whole of the data from each preparation. This method can be used whatever the intervals between doses, even if they are not the same for the two preparations. The residual sum of squares between doses, after subtraction of the 'Preparations' and 'Regression' components, cannot easily be subdivided, but its mean square may be compared with the error mean square (with the aid of Appendix II) to give a composite test of non-parallelism and non-linearity. This test corresponds to that which would be obtained if the last three of the single degrees of freedom in Table XIX were added together; it is not able to distinguish different sources of invalidity, but, providing neither the test nor the response diagram suggest invalidity, calculation of M, R, and the fiducial limits may proceed exactly as in § 58. A numerical example has been given in a publication of the British Standards Institution (1940); see also Emmens (1948).

Considerations of optimal design (Section XI), as well as the convenience of statistical analysis, make symmetrical assays preferable wherever they are possible.

VII. Slope-ratio Assays

60. Microbiological assays. Not all methods of biological standardization depend upon a response which has a linear regression on the logarithm of the dose. Many microbiological assay techniques, in particular, give responses which have a linear regression on the dose itself. Estimation of the regression from data relating to a single preparation presents no difficulties: an example has been given in §§ 30–5. When simultaneous tests of two preparations, S and T, are made, however, the situation is different from that discussed in Section VI. The two regression lines will not be parallel, but will intersect at a point for which the dose is zero. If the region of linearity of the dose-response law extends to zero dose, as often seems to be

true, the point of intersection will represent the response to zero dose; if not, the point of intersection will probably be lower than the true response to zero dose, but will still occur at this level of dose. A statistical technique different from that of Section VI is therefore required.

Though this form of relationship seems most common with microbiological assay techniques, there are undoubtedly some macrobiological assays in which it is needed.

61. Data for 5-point common-zero assay. The responses in Table XXI were obtained in an assay of malt for its riboflavin content. *Lactobacillus helveticus* was the test organism, and its growth was assessed by titration with N/10 NaOH (Wood, 1946). Four tubes with single and double doses of a standard riboflavin preparation and of a preparation containing the malt were included, as well as four control or 'blank' tubes. Inspection of Fig. 13, in which different scales have been used for units of the two preparations, in order to keep the lines distinct, makes clear that the pairs of points for S and T lie approximately in line with the point for zero dose.

TABLE XXI. *Growth of* L. helveticus (*in ml.* N/10 NaOH) *for various doses of riboflavin or malt*

Standard preparation (x_S) . . 1 unit = 0·1γ riboflavin.
Test preparation (x_T) . . 1 unit = 0·025 gm. malt.

	Blanks	*Standard*		*Test*	
	$x_S = 0$ $x_T = 0$	$x_S = 1$ $x_T = 0$	$x_S = 2$ $x_T = 0$	$x_S = 0$ $x_T = 1$	$x_S = 0$ $x_T = 2$
	1·9	4·8	8·4	4·0	6·0
	2·2	5·0	8·2	4·4	6·2
	2·0	5·2	8·0	4·5	6·1
	2·2	4·9	7·8	4·1	6·1
Totals	8·3	19·9	32·4	17·0	24·4
Means	2·08	4·98	8·10	4·25	6·10

A rapid assessment of potency could be made from Fig. 13 with very little calculation. Two lines might be drawn by eye, subject to the condition that they should intersect at $x = 0$, though not necessarily having exactly 2·08 as the response at their intersection, for that response is as subject to error as are those for the non-zero doses. The lines would show for the

standard preparation an increase in response of almost exactly 3 per 1 unit (or 0·1γ) increase in dose and for the test preparation an increase in response of almost exactly 2 per 1 unit

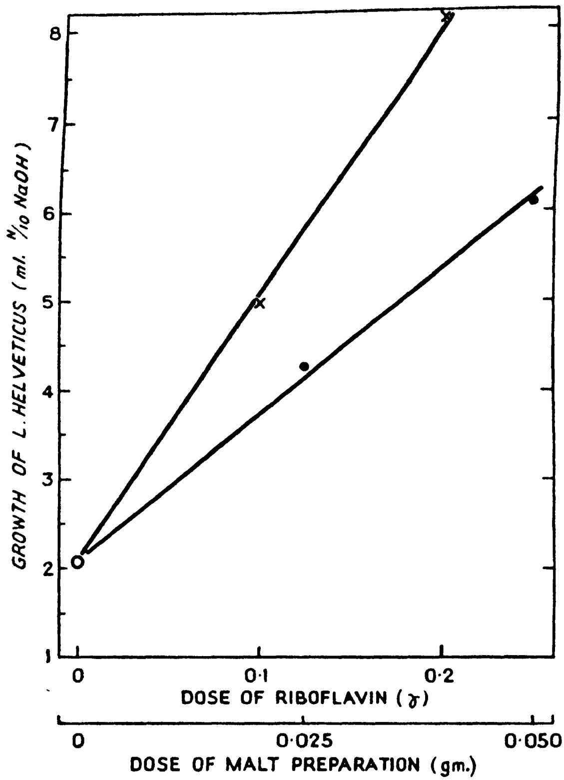


FIG. 13. Assay of riboflavin by growth of *Lactobacillus helveticus*. (Data in Table XXI.)

× Standard preparation. ● Test preparation.

(0·025 gm.) increase in dose. Hence the relative potency, the ratio of equally effective doses, is estimated to be

$$R = \tfrac{2}{3};$$

this represents the riboflavin content, in units of 0·1γ, of 0·025 gm. malt, and therefore 1 gm. malt is estimated to contain

$$40 \times \tfrac{2}{3} \times 0{\cdot}1\gamma = 2{\cdot}7\gamma \text{ riboflavin.}$$

Once again arithmetical analysis is required if the validity of the assay and the precision of the estimation are to be examined.

62.† Multiple regression. The requirement in an assay of this kind is to fit two straight lines, one to each preparation, subject to the condition that they must intersect at zero dose. The equations for the two lines may be written

$$\left.\begin{aligned} Y &= a+b_S x \\ \text{and} \qquad Y &= a+b_T x, \end{aligned}\right\} \tag{36}$$

where x is the dose of either preparation; here the constant term in the equations is the same for both preparations, whereas in parallel line assays the regression coefficient was the same for both equations. For the purpose of computation, the procedure may be regarded as fitting a *multiple regression equation*, an equation stating the average response as dependent upon two different x-variates, to all the data. In fact, the two equations already written may be combined into

$$Y = a+b_S x_S+b_T x_T, \tag{37}$$

where x_S (the dose of the standard preparation) is zero for the blanks and for all doses of the test preparation, and x_T (the dose of the test preparation) is zero for the blanks and all doses of the standard. The calculations for the regression coefficients are described in books on statistical method (Fisher, 1946*a*, § 29; Mather, 1946, § 38). They necessitate the solution of the pair of linear equations

$$\left.\begin{aligned} b_S S_{x_S x_S}+b_T S_{x_S x_T} &= S_{x_S y} \\ b_S S_{x_S x_T}+b_T S_{x_T x_T} &= S_{x_T y}, \end{aligned}\right\} \tag{38}$$

where each of the 'S' symbols represents a sum of squares or sum of products of deviations formed according to equations (15) and (16). For example,

$$\begin{aligned} S_{x_S x_S} &= Sx_S^2-\frac{(Sx_S)^2}{N} \\ &= 4\times(1^2+2^2)-\frac{(4\times 1+4\times 2)^2}{20} \\ &= 12{\cdot}8, \\ S_{x_S x_T} &= Sx_S x_T-\frac{(Sx_S)(Sx_T)}{N} \\ &= 0-\frac{(4\times 1+4\times 2)(4\times 1+4\times 2)}{20} \\ &= -7{\cdot}2, \end{aligned}$$

and

$$S_{x_S y} = Sx_S y - \frac{(Sx_S)(Sy)}{N}$$
$$= 1 \times 19\cdot9 + 2 \times 32\cdot4 - \frac{(4 \times 1 + 4 \times 2)(8\cdot3 + 19\cdot9 + 32\cdot4 + 17\cdot0 + 24\cdot4)}{20}$$
$$= 84\cdot7 - 61\cdot2$$
$$= 23\cdot5.$$

Because of the symmetry of the assay, $S_{x_T x_T}$ and $S_{x_S x_S}$ are equal, but the method of analysis applies even when this is not so. Equations (38) take the form

$$12\cdot8 b_S - 7\cdot2 b_T = 23\cdot5,$$
$$-7\cdot2 b_S + 12\cdot8 b_T = 4\cdot6.$$

By elementary algebra, the solutions are

$$b_S = 2\cdot982,$$
$$b_T = 2\cdot037.$$

The regression equation, analogous to equation (13), has the form

$$Y = \bar{y} + b_S(x_S - \bar{x}_S) + b_T(x_T - \bar{x}_T), \qquad (39)$$

which corresponds to equation (37), with

$$a = \bar{y} - b_S \bar{x}_S - b_T \bar{x}_T.$$

The means $\bar{x}_S$, $\bar{x}_T$, $\bar{y}$ relate to all the 20 tubes, so that

$$\bar{x}_S = \bar{x}_T = \frac{4 \times 1 + 4 \times 2}{20}$$
$$= 0\cdot6,$$
$$\bar{y} = \frac{8\cdot3 + 19\cdot9 + 32\cdot4 + 17\cdot0 + 24\cdot4}{20}$$
$$= 5\cdot1.$$

Equation (39) may therefore be written

$$Y = 2\cdot09 + 2\cdot98 x_S + 2\cdot04 x_T,$$

an equation which, by taking first $x_T = 0$ and then $x_S = 0$, gives equations (36) as

$$\left.\begin{aligned} Y &= 2\cdot09 + 2\cdot98 x_S \\ \text{and} \quad Y &= 2\cdot09 + 2\cdot04 x_T, \end{aligned}\right\}$$

the two lines plotted in Fig. 13. Though the diagram is not needed for the relative potency calculation, completion of it is a valuable safeguard against gross errors of arithmetic, as inspection ensures that the lines are, at least approximately, in the right places.

63.† Analysis of variance. Once again an analysis of variance provides an estimate of variance per subject and an important test of assay validity. The total, between doses, and error components in Table XXII are obtained in the usual manner. That for regression, according to standard procedure for multiple regression analysis, has a sum of squares

$$b_S S_{x_S y} + b_T S_{x_T y} = 2{\cdot}982 \times 23{\cdot}5 + 2{\cdot}037 \times 4{\cdot}6 \qquad (40)$$
$$= 79{\cdot}447,$$

with 2 degrees of freedom since two regression coefficients have been estimated.[1] A sum of squares for 'deviations from regression' is then obtained by subtraction from the sum of squares between doses. The error mean square, as usual, is to be taken as s^2:

$$s^2 = 0{\cdot}0363.$$

The analysis can easily be adapted for assays with several doses of each preparation (not necessarily equal numbers), there being then more than 2 degrees of freedom for 'deviations from regression'.

TABLE XXII. *Analysis of variance for the data of Table XXI*

Adjustment for mean . . .		520·200	
Nature of variation	*d.f.*	*Sum of squares*	*Mean square*
Regression	2	79·447	
Deviations from regression .	2	0·108	0·059
Between doses	4	79·555	
Error	15	0·545	0·0363
Total	19	80·100	

64. Validity tests. Further analysis of the sum of squares for deviations from regression is not easy, but a comprehensive test of significance against the error mean square may be made.

[1] Note that for a simple regression, the sum of squares for the regression component, usually written $(S_{xy})^2/S_{xx}$, is equal to bS_{xy}.

A significant mean square for deviations would indicate either that the assay was invalid or that treatment of the results by the slope-ratio method was inappropriate. For example, the regressions might not be linear, so that an analysis which employed linear regression on dose would be unsuitable, or the regressions, though linear, might fail to intersect at $x = 0$, so contradicting the whole basis of the assay—the hypothesis of similarity; inspection of Fig. 13 would help to indicate the source of trouble. Here the ratio of mean squares

$$F = \frac{0{\cdot}059}{0{\cdot}036} = 1{\cdot}6,$$

with $n_1 = 2$, $n_2 = 15$ is well below the significance level, 3·7, shown in Appendix III, and no difficulty arises. Estimation of the relative potency may proceed.

65. Relative potency. Each regression coefficient represents the effect on response of unit increase in dose. Hence unit dose of the test preparation is equal in effect to R units of the standard, where

$$R = \frac{b_T}{b_S} \tag{41}$$

$$= 0{\cdot}683.$$

From the definition of the units of the two preparations, at the head of Table XXI, 0·025 gm. malt is estimated to contain $0{\cdot}1 \times 0{\cdot}683\gamma$ riboflavin, so that 1 gm. malt contains $4 \times 0{\cdot}683$ or $2{\cdot}73\gamma$ riboflavin (cf. § 61).

66.† Fiducial limits. The relative potency estimate contains b_S in its denominator, and therefore an approximate variance of R can be used safely only if b_S is much greater than its standard error. Clearly, if the experiment failed to provide significant evidence that the response was dependent upon the dose for the standard preparation, any attempt at assaying T in terms of S would be useless. Once again a g-criterion can be applied as a test of whether an approximate variance can be used, and, though the expression for g looks very different from that of equation (34), it performs the same function. If

$$g = \frac{s^2 t^2 S_{x_T x_T}}{b_S^2\{S_{x_S x_S} S_{x_T x_T} - (S_{x_S x_T})^2\}} \tag{42}$$

is small (say, less than 0·1) the approximate variance may be used, otherwise an exact formula for fiducial limits (Finney, 1947*b*) is required; a value of g greater than unity would indicate that there was no significant regression of response on dose, and the assay would have to be abandoned. In fact, for good microbiological assays, g will always be negligible (Wood and Finney, 1946), though other slope-ratio assays need not have this property. Here

$$g = \frac{0\cdot0363 \times (2\cdot13)^2 \times 12\cdot8}{(2\cdot982)^2 \times \{(12\cdot8)^2 - (7\cdot2)^2\}}$$
$$= 0\cdot002.$$

The expression for $V(R)$ is complicated in appearance, but not difficult to use:

$$V(R) = \frac{s^2}{b_S^2} \times \frac{S_{x_S x_S} + 2RS_{x_S x_T} + R^2 S_{x_T x_T}}{S_{x_S x_S} S_{x_T x_T} - (S_{x_S x_T})^2} \tag{43}$$
$$= \frac{0\cdot0363}{(2\cdot982)^2} \times \frac{12\cdot8 - 14\cdot4 \times 0\cdot683 + 12\cdot8 \times (0\cdot683)^2}{(12\cdot8)^2 - (7\cdot2)^2}$$
$$= 0\cdot000326.$$

Hence $$R = 0\cdot683 \pm 0\cdot018,$$

and the fiducial limits, $2\cdot13 \times 0\cdot018$ on either side of R, become 0·645, 0·721. The estimated potency, 2·73γ riboflavin per gm. malt, is therefore assigned fiducial limits of 2·58γ, 2·88γ per gm. malt.

67.† Response metameters. The parallel line assay, using a linear regression of response on log dose, and the slope-ratio assay, using a linear regression of response on dose, are the types of chief importance to-day for quantitative data. They are not the only types, however. Transformation of response to a *response metameter* (e.g. use of the logarithm, or the square root of the measured response) may sometimes be needed, in order to provide a linear relationship amenable to statistical analysis. Wood (1947) has shown that for microbiological assays of many amino-acids the logarithm of response has a linear regression on log dose; the method of analysis then is essentially the same as for the parallel line assays already discussed, but transformation to a response metameter introduces certain difficulties in

connexion with the variance of responses, the theory of which has been discussed by Fieller (1947) and Finney (1947*b*).

VIII. The Binomial and χ^2 Distributions

68. Quantal data. The results of biological experiments and tests often have to be expressed qualitatively rather than quantitatively, that is to say, by classification of individual records into two or more groups instead of by measurement of each. Qualitative classification necessarily conveys less exact information, but the obtaining of a measurement for every individual may be impossible or exceedingly laborious, so rendering the cruder form of record unavoidable. Particularly important in connexion with biological standardization is the statistical treatment of dichotomous classifications, providing estimates of the percentages of subjects which show a certain response under specified conditions. Many methods of standardization, for example, involve comparisons between percentage mortalities in groups of animals receiving various doses of standard and test preparations. Death, indeed, is perhaps the *quantal response* most commonly used in standardization techniques, but it is by no means the only one; the mouse convulsion test for insulin assay (Chapter V) involves comparisons of percentages of animals convulsed by different doses, and oestrone may be assayed by means of comparisons between the percentages of rats in which oestrus is produced (Chapter IX).

69. Binomial distribution. Before studying the special statistical problems of biological assays which use quantal responses, the simpler statistical properties of data relating to qualitative classifications must be briefly discussed. Suppose that, in a test of the toxicity of an arsenical compound to mice, a dose is given that is just sufficient (though the experimenter cannot know this) to cause the death of 50 per cent. of the population of mice from which those used in the test were selected. It will not necessarily kill exactly 50 per cent. of the sample of mice tested, and replicate tests, using several samples, will almost certainly give different results. As an extreme case, if the 'sample' consisted of 1 mouse, either of two events could occur: either 0 or 1 would die, and, since the two results are equally likely, each has a probability of 0·5. For a sample of 2 mice, the possible

results are AA, DA, AD, DD, where the letters A and D represent 'alive' and 'dead', and the ordered pairs of letters give the records for the first and second mouse. If the effect of the poison on either mouse is independent of that on the other, these four possibilities will all have the same probability, 1/4. Both DA and AD represent the same death-rate, 1 out of 2, and the total probabilities for 0, 1, and 2 deaths are therefore 0·25, 0·50, and 0·25 respectively. In other words, if a large number of samples of 2 are tested, though half of them will show 1 death, or '50 per cent. kill', one-quarter will show no deaths and one-quarter 2 deaths. Similarly, in a sample of 3 mice, the 8 possibilities AAA, DAA, ADA, DDA, AAD, DAD, ADD, and DDD each have the probability 1/8; addition shows the probabilities of 0, 1, 2, and 3 deaths to be 0·125, 0·375, 0·375, and 0·125 respectively.

In a sample of n mice, each mouse must fall in one of the two categories A and D, and the total number of possible sequences of results, such as $DDAD...ADA$, is 2^n. If the responses are all independent of one another, these must all have the same probability, $1/2^n$. The investigator usually has little interest in distinguishing the individual mice, and wishes to record only whether the number of deaths is 0, 1, 2, 3,..., $(n-1)$, or n. The number of sequences of results in which there are exactly r deaths is the number of possible arrangements of n letters such that r of them are D and $(n-r)$ are A, the total of which is known, from the algebraic theory of permutations and combinations, to be

$$\frac{n!}{r!(n-r)!}. \qquad (44)$$

The quantity denoted by $r!$, and read as 'factorial r', is to be understood as the product $r(r-1)(r-2)...\times 3\times 2\times 1$, so that, for example,

$$7! = 7\times 6\times 5\times 4\times 3\times 2\times 1 = 5{,}040;$$

the symbol $0!$ is always to be interpreted as 1 (the same value as $1!$) for reasons of mathematical theory. Consequently, the total probability that exactly r deaths occur, without regard to which mice die, is

$$\frac{n!}{r!\,(n-r)!}\times (\tfrac{1}{2})^n.$$

Evaluation of this expression for selected values of n and r will show that numbers of deaths near to $\frac{1}{2}n$ always have the largest probabilities, and that, except when n is small, the probability that almost all survive or almost all die is very small. For $n = 10$, the probability that all survive is 0·001, as also is the probability that all die, the probabilities that 1 or 9 die are each 0·011, so that in all but a proportion 0·024 of samples the number of deaths will lie between 2 and 8.

A generalization of this result is needed if the dose tested is sufficient to kill a proportion P of the population of mice, where P is not necessarily 0·5. The 8 possible sequences of results from a sample of 3 no longer have equal probabilities; instead the probabilities are $(1-P)^3$, $P(1-P)^2$, $(1-P)P(1-P)$, $P^2(1-P)$, $(1-P)^2P$, $P(1-P)P$, $(1-P)P^2$, and P^3 respectively. The general theorem is that the probability of exactly r deaths in a sample of n is

$$\frac{n!}{r!\,(n-r)!}P^rQ^{n-r}, \tag{45}$$

where Q is written for $(1-P)$. This probability is the $(r+1)$th term of the binomial expansion of $(Q+P)^n$, and the set of probabilities of the different numbers of deaths out of a total of n is known as the *binomial probability distribution*. In a long series of samples, the average number of deaths per sample will be close to nP; in any one sample, however, all numbers from 0 to n must be regarded as possible, though those in the neighbourhood of nP are much more probable than extreme values. For example, if 5 mice are injected with a dose which on an average kills 80 per cent., the probability that exactly 4 will die is given by formula (45) as

$$\frac{5!}{4!\,1!}\times(0{\cdot}8)^4\times(0{\cdot}2) = \frac{5\times4^4}{5^5}$$

$$= 0{\cdot}4096.$$

Similarly the probability that exactly 3 will die is

$$\frac{5!}{3!\,2!}\times(0{\cdot}8)^3\times(0{\cdot}2)^2 = \frac{10\times4^3}{5^5}$$

$$= 0{\cdot}2048.$$

Evaluation for other values of r shows that the probabilities of

0, 1, 2, 3, 4, and 5 deaths are respectively 0·0003, 0·0064, 0·0512, 0·2048, 0·4096, and 0·3277.

70. Probabilities of deaths. Appendix IV (Durham, Gaddum, and Marchal, 1929) shows numerical values of the probabilities that r or less deaths occur in a sample of n. The five parts of the table relate to $n = 5$, 10, 15, 20, and 30, for each of which the proportion P is assigned in turn the values 0·025, 0·05, 0·075, 0·10, 0·15, 0·20,..., 0·85, 0·90, 0·925, 0·95, 0·975. The probabilities for particular values of r were calculated by formula (45), after which the probability of 0 deaths was added to the probability of 1 death, this sum to the probability of 2 deaths, and so on. The mode of construction will be understood by comparing the probabilities for $P = 0{\cdot}8$, $n = 5$, shown at the end of § 69, with the 80 per cent. column in Appendix IV (*a*). The tables may be of value to those engaged in framing tests in which a percentage response, whether death or other physiological change, is to be observed in a small sample of animals.

71.† Limits of error. Appendix IV may also be used in an inverse manner to give limits of error for a mortality rate. Suppose that in a batch of 20 mice 7 deaths occur. The mortality rate to be expected in the population of animals from which the batch was selected, if all were given the same dose, will then be estimated as 7/20 or 35 per cent. Clearly this need not be the true rate. If the true rate were, say, 25 per cent., the probability of finding as many as 7 deaths in a sample of 20 would still be quite high; similarly, if the true rate were 45 per cent. the probability of as few as 7 deaths in a sample of 20 would be quite high. The fiducial argument introduced in § 14 is strictly applicable only to measurements having a continuous distribution. Nevertheless, limits of error, having a logical basis very similar to that of fiducial limits, can be defined: the lower of these limits is the lowest value of P which could reasonably be expected to give as many as 7 deaths, and the upper is the highest value of P which could reasonably be expected to give as few as 7 deaths. From inspection of Appendix IV (*d*), if P were about 15 per cent., the probability of 6 deaths or less would be 0·975 and, therefore, the probability of 7 deaths or more would be 0·025. Similarly, if P were a little less than 60 per cent. (say, 59 per cent.), the probability of 7 deaths or less would be 0·025. Consequently the mortality rate

may be asserted to lie between 15 per cent. and 59 per cent., and the probability that statements made in this way will be wrong is only 0·05. Numerical values of these limits can be derived more accurately from Stevens's table (Fisher and Yates, 1948, Table VIII_1), and are 15·4 per cent., 59·2 per cent.

72.† Small samples. Trevan (1927) has emphasized the risk of drawing false conclusions from the uncritical examination of data from small samples. For example, if two toxic substances are tested, each on a small batch of animals, and the true potencies differ considerably, the experimental records may by chance show no difference or even a difference in the wrong direction. If the true mortality rates were 40 per cent. and 80 per cent., in batches of 5 animals on an average 2 and 4 would die respectively. In any one test, however, there are six possible results for each preparation, calculated from formula (45) as:

	Probability of results for	
No. of deaths	$P = 0{\cdot}4$	$P = 0{\cdot}8$
0	0·0778	0·0003
1	0·2592	0·0064
2	0·3456	0·0512
3	0·2304	0·2048
4	0·0768	0·4096
5	0·0102	0·3277
Total . .	1·0000	1·0000

Simple calculation shows that the probability of equal numbers of deaths in any one experiment, the sum of products of corresponding entries for $P = 0{\cdot}4$ and $P = 0{\cdot}8$, is 0·1014. Similarly, the probability that the less potent preparation will give more deaths than the other is

$$\begin{aligned}
&0{\cdot}2592 \times 0{\cdot}0003 \\
&\quad + 0{\cdot}3456 \times (0{\cdot}0003 + 0{\cdot}0064) \\
&\quad + 0{\cdot}2304 \times (0{\cdot}0003 + 0{\cdot}0064 + 0{\cdot}0512) \\
&\quad + 0{\cdot}0768 \times (0{\cdot}0003 + 0{\cdot}0064 + 0{\cdot}0512 + 0{\cdot}2048) \\
&\quad + 0{\cdot}0102 \times (0{\cdot}0003 + 0{\cdot}0064 + 0{\cdot}0512 + 0{\cdot}2048 + 0{\cdot}4096) \\
&= 0{\cdot}0428.
\end{aligned}$$

Thus in 10·1 per cent. of trials the two preparations will appear

to be equipotent, and in 4·3 per cent. the preparation for which $P = 0{\cdot}8$ will appear to be less potent than that for which $P = 0{\cdot}4$; only 85·6 per cent. of trials will correctly indicate the more potent preparation. An increase in the size of the batches would reduce the frequency with which the potencies appeared to be placed in the wrong order, but could not ensure that each would kill exactly the right proportion.

73. Normal approximation. For larger samples, the behaviour of the binomial distribution is satisfactorily approximated by a normal distribution. If, in a batch of n animals, r show the characteristic response,

$$p = r/n \tag{46}$$

is an estimate of P, the true response rate for the population. The variance of this estimate can be proved to be

$$V(p) = PQ/n, \tag{47}$$

and p may be assigned a standard error calculated as the square root of this expression. When P is unknown, as in practice it usually is, the variance itself must be estimated as

$$V(p) = pq/(n-1), \tag{48}$$

where q is written for $(1-p)$. These results are true for any value of n, but they are of limited use unless n is large (say, at least, 100). For large n fiducial limits may be assigned on the approximation that p is normally distributed, using a range 1·96 times the standard error on either side of p for the 0·05 limits. For example, if 82 out of a batch of 230 mice are killed, from equation (46)

$$p = 82/230$$
$$= 0{\cdot}357$$

is an estimate of the true mortality rate, and from equation (48) the variance is estimated as

$$V(p) = (0{\cdot}357 \times 0{\cdot}643)/229$$
$$= 0{\cdot}0010.$$

Hence the standard error of p is 0·032. Multiplication by 1·96 gives 0·063 as the distance of the fiducial limits on either side of p, and therefore P may be asserted to lie between 0·294 and 0·420.

In all this argument, it is most important that the animals

should react to the treatment independently of one another. If, in the example just quoted, the 230 mice consisted of 5 from each of 46 litters, such that 16 litters were peculiarly susceptible to the poison and all died, whereas the other 30 litters were resistant and lost only 2 mice, clearly the effective number of independent observations would be about 46 rather than 230, and the precision of p would be much reduced.

74. Significance of a difference. The argument used by Trevan has shown the danger of trusting superficial indications in mortality or other quantal response data. In small batches of subjects, even large true differences in potency may escape detection or appear to have their effects reversed. The data in Table XXIII (Irwin, 1943) were obtained in a comparison of two insulin preparations by the mouse-convulsion test. The proportions showing convulsions in the two groups are 0·497 and 0·567, but this might not indicate a genuinely higher potency for the test preparation. As for data involving measurement, a satisfactory judgement on the reality of the potency difference can be made only after application of a test of significance.

TABLE XXIII. *Comparison of insulin preparations by mouse-convulsion test*

	Observed frequencies			*Expected frequencies*	
	Standard preparation	*Test preparation*	*Totals*	*Standard preparation*	*Test preparation*
Mice not convulsed . .	151	65	216	144·0	72·0
Mice convulsed .	149	85	234	156·0	78·0
Totals . .	300	150	450	300	150

75. Expected frequencies. If the true potencies were in fact equal, the best estimate of the convulsion rate would be obtained from the totals of the two preparations

$$p = 234/450$$
$$= 0{\cdot}520.$$

This proportion may be used to calculate *expected frequencies* of convulsed and not convulsed for each group, as shown in the

right-hand half of the table, on the hypothesis of equipotency: for the standard

$$300 \times 0{\cdot}520 = 156{\cdot}0,$$

$$300 \times 0{\cdot}480 = 144{\cdot}0.$$

By chance, these expected frequencies work out exactly to whole numbers, but they need not always do so. Comparisons between the two parts of the table show that more mice were convulsed on the test preparation, less on the standard preparation, than is predicted on the hypothesis of equipotency. The discrepancy between observation and expectation in each of the four 'cells' of the table is 7·0, and a test of significance may be based upon this quantity.

76. The χ^2 test. The procedure is to reduce the discrepancy by 0·5, square it, and form the sum of the results of dividing the square by each of the expected frequencies in turn. This quantity, known as χ^2, may be written:

$$\chi^2 = (\text{Discrepancy} - 0{\cdot}5)^2 \times \text{Sum of reciprocals of expected frequencies.} \qquad (49)$$

Clearly the value of χ^2 will increase if the discrepancy be increased. Its distribution, on the hypothesis of no real difference in potency, is the χ^2 *distribution* (Appendix V) with 1 degree of freedom. Table XXIII and equation (49) give

$$\chi^2 = (6{\cdot}5)^2 \times \left(\frac{1}{144{\cdot}0} + \frac{1}{156{\cdot}0} + \frac{1}{72{\cdot}0} + \frac{1}{78{\cdot}0}\right)$$
$$= 1{\cdot}7.$$

From the first line of Appendix V, it is seen that values of χ^2 as great as 3·8 could occur by chance with a probability 0·05, and a value of 1·7 cannot be regarded as significant evidence that the potencies of the two preparations are in reality unequal. Had the test preparation given 90 convulsed mice out of 150, similar calculations would have led to

$$\chi^2 = 3{\cdot}9,$$

a value just large enough to be judged significant; the evidence would then have sufficed to demonstrate a real difference in potency.

The method of calculation of χ^2 described above is important because of its easy generalization to more complex problems.

An alternative for this simple type of table, leading to exactly the same result, is

$$\chi^2 = \frac{450 \times (151 \times 85 - 149 \times 65 - 225)^2}{300 \times 150 \times 216 \times 234}$$
$$= 1{\cdot}7. \qquad (50)$$

The first term in the numerator is the total number of mice. Within the brackets is a difference of 'cross-products' of the four observed frequencies in Table XXIII, which must always be taken as positive, from which is subtracted 225, one-half of the total number of mice. The denominator is the product of the four marginal totals in Table XXIII. The adjustment of the discrepancy by 0·5 in the first form of calculation, which has the same effect as the subtraction of 225 from the difference of cross-products in the second, is known as the *Yates correction for continuity*: it is often omitted when the number of observations is very great, since its effect is then trivial, but since its inclusion is so simple the general rule should be to include it. When the number of observations is very small, and in particular when any one of the expected frequencies is less than 5·0, the test described here is untrustworthy. Fisher (1946*a*, § 21.02) describes the method then required, and Finney (1948*b*) gives a table to assist its use.

77. The $2 \times c$ contingency table. The χ^2 test may be applied more generally, in order to test the significance of deviations from proportionality of frequencies in *contingency tables* more extensive than that of two rows and two columns given above as Table XXIII. An important use is for contingency tables analogous to the 2×2, but with c columns instead of only two. The figures in Table XXIV relate to an investigation of the vaginal smear method proposed for the assay of gonadotropic hormone. Four doses were tested on comparable groups of rats, and the numbers of animals in which vaginal cornification occurred was observed. The numbers of non-responding and responding rats at each dose are shown in the first and third lines of Table XXIV; immediately below, in the second and fourth lines, are shown corresponding expected frequencies, whose calculation is about to be described. The fifth line gives the total number of rats (observed or expected) at each dose. The observed percentage of animals responding, shown in the

last line, ranged from 44 at the lowest dose to 78 at the highest, but between these varied erratically; calculation of χ^2 leads to a test of whether the difference can be attributed to chance.

TABLE XXIV. *Vaginal cornification of rats in response to gonadotropic hormone*

	Dose in mg.				
	0·3	0·6	1·2	2·4	*Totals*
No response, observed	10	7	10	5	32
No response, expected	6·6	8·5	8·5	8·5	
Response, observed .	8	16	13	18	55
Response, expected .	11·4	14·5	14·5	14·5	
Totals . . .	18	23	23	23	87
Percentage responding	44	70	57	78	63·2

If there were no effect of dose on the proportion of rats responding, that proportion would be estimated from the totals for all doses as

$$p = 55/87$$
$$= 0{\cdot}632.$$

This value can be used to calculate 'expected frequencies' in each of the eight cells of the table; for example, the expected number failing to respond at a dose of 1·2 mg. is $23\times(1-0{\cdot}632)$, or 8·5. The discrepancy between each expected frequency and the corresponding observation is then squared, divided by the expectation, and the results summed to give:

$$\chi^2 = S\frac{(\text{observed}-\text{expected})^2}{\text{expected}} \tag{51}$$

$$= (3{\cdot}4)^2\times\left(\frac{1}{6{\cdot}6}+\frac{1}{11{\cdot}4}\right)+(1{\cdot}5)^2\times\left(\frac{1}{8{\cdot}5}+\frac{1}{14{\cdot}5}+\frac{1}{8{\cdot}5}+\frac{1}{14{\cdot}5}\right)$$
$$+(3{\cdot}5)^2\times\left(\frac{1}{8{\cdot}5}+\frac{1}{14{\cdot}5}\right)$$
$$= 5{\cdot}9.$$

Note that the Yates correction is not used here; it is applicable only to 2×2 contingency tables, such as that in Table XXIII. Except for this modification, equations (49) and (51) are identical, since under the conditions in which (49) is used the

discrepancy between observed and expected frequencies is the same for all the four cells of the table. The χ^2 obtained has $(c-1)$ degrees of freedom, 1 less than the number of columns, and its significance is tested with the aid of the appropriate line of Appendix V. For 3 degrees of freedom, the 0·05 significance level is 7·8. Hence the χ^2 for Table XXIV does not indicate any serious deviation from the hypothesis of a response rate independent of dose. An experiment on a larger scale would be sensitive to smaller differences, and might disclose some effect of dose, but it seems clear that, unless considerably more than 100 rats can be spared for each preparation under examination, variations in the response rate will be of little use as a basis for biological assay of the hormone.

Data such as those of Table XXIV, in which the columns of the contingency table represent different doses of a drug, might also be examined by the probit regression technique described in Section IX. Indeed, that method is generally preferable because it concentrates attention on the linear trend, the single and most important degree of freedom, instead of on all differences between columns. The χ^2 test, however, can be applied more quickly, and is useful here as confirmation of what seems obvious by inspection of Table XXIV, namely, that the response rate does not increase with dose. The χ^2 test would still be applicable if the columns represented not several doses of one drug, but doses of different drugs, or a fixed dose of one drug injected by different methods—in other words, if differences between the treatments shown in separate columns were qualitative instead of quantitative. The roles of rows and columns of a contingency table can, of course, be interchanged, and the same χ^2 test may be used for data arranged as a contingency table of r rows and 2 columns.

An alternative method of calculation of χ^2 for the $2 \times c$ table, advantageous because of its speed, has been described by Fisher (1946*a*, § 21), Mather (1946, § 50), and Snedecor (1946, § 9.10). It is not shown here, as it does not generalize for use with the $r \times c$ table.

78.† The $r \times c$ contingency table. The same statistical test can be used in the examination of discrepancies from proportionality in contingency tables having more than two rows. For example, in an experiment such as that whose results are

recorded in Table XXIV, the animals receiving any one dose might be classified into several groups according to the nature of an observed response, instead of by a simple dichotomy. The classification might be according to intensity of response on an arbitrarily graded scale, such as the assessment of degree of healing in the line test for vitamin D (§ 22), though analysis of variance of the results would then probably be the most convenient method of comparing different doses or different preparations. The merit of the χ^2 method is that it applies equally well when both rows and columns represent purely qualitative classifications. The columns might represent doses of different preparations, the rows frequencies of various types of response, not necessarily with any known order of increasing intensity. The χ^2 test can still be used to examine the proportionality of the various responses for different preparations; it takes no account of the order in which either rows or columns are arranged, and so neglects some information when (as in Table XXIV) one or both categories can be assigned measures on a numerical scale.

The calculation of χ^2 is just as in § 77. Expected frequencies in every cell of the table, on the hypothesis of proportionality, are first found. The simplest rule is to form totals by rows and by columns, and to take as the expected frequency in any cell the product of the corresponding row and column totals divided by the total number of observations: in Table XXIV, for the expected frequency of responses at the lowest dose

$$(55 \times 18)/87 = 11{\cdot}4.$$

If there are r rows and c columns, there will be rc discrepancies between observed and expected frequencies; χ^2 is found from equation (51) as a sum of rc terms, and must be assigned $(r-1)(c-1)$ degrees of freedom. The Yates correction is to be used only when $r = c = 2$, but for all values of r and c a χ^2 calculated from data giving one or more expected frequencies less than 5 may be untrustworthy.

No example of these calculations will be given, as, important though the method is for many statistical purposes, it is seldom required in standardization problems. Further illustrations of the use of χ^2 may be found in the books by Fisher, Mather, and Snedecor already mentioned.

The reader must remember that the χ^2 test described here is applicable only to the statistical analysis of classified frequency tables, and not to percentages or to sets of average frequencies, except by reconversion of these to true observed frequencies. *The test must in no circumstances be used for measurements of weight, length, time, etc.*

IX. The Probit Transformation

79. Tolerance distribution. If a quantal response is to be made the basis of a biological assay technique, the relationship between the dose of the material to be assayed and the frequency with which that dose produces the response must be studied. The investigation summarized in Table XXIV gave no suggestion that the response rate depended upon dose; in other circumstances, groups of subjects which are given various doses of a drug or other agent will show a steady increase (or a steady decrease) in the proportion manifesting the characteristic response to be associated with increase in dose. The results shown in Table XXV are typical. Trevan (1927) records them as obtained in tests of the toxicity of various concentrations of cocaine hydrochloride to mice. The percentage mortality is plotted against log dose in Fig. 14. A simple linear regression of percentage on log dose could scarcely be appropriate for data of this kind, since all responses must lie between 0 per cent. and 100 per cent.

Table XXV. *Toxicity of cocaine hydrochloride (0·1 per cent. solution) to mice*

Dose (mg. per gm. body-weight)	*Number of mice*		*Percentage kill*
	Injected	*Killed*	
0·015	20	0	0
0·020	69	11	16
0·025	95	50	53
0·030	78	61	78
0·035	44	37	84
0·040	20	20	100

A better understanding of the problem will be obtained by comparison with the type of record that results from direct measurement of lethal dose. The strophanthus data in Table IX, for example, were obtained by slow infusion into cats,

which was stopped as soon as death occurred, so providing a direct measure of the amount needed to kill each cat. The results give information on the distribution of tolerances, or just-lethal doses, for the population from which the cats were

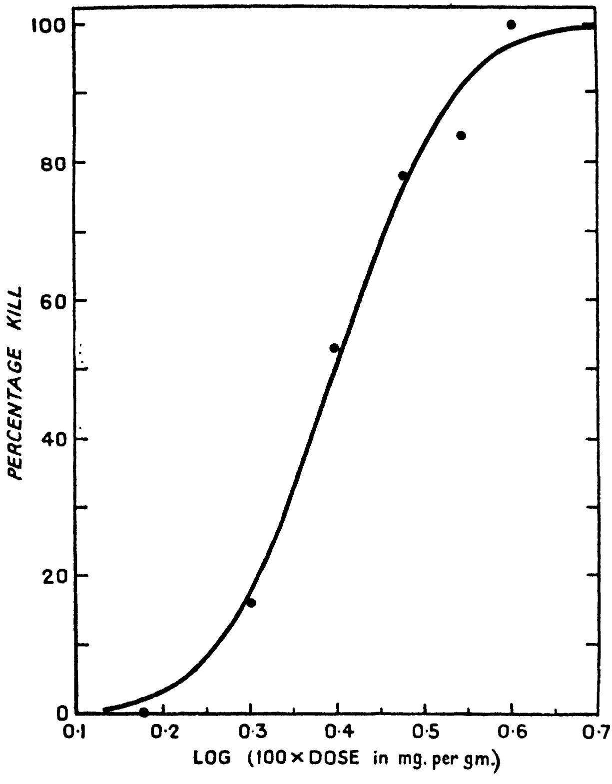


FIG. 14. Toxicity of cocaine hydrochloride to mice. (Data in Table XXV.) The sigmoid curve is obtained from the calculations described in §§ 84–92.

selected. On the other hand, for the data of Table XXV no measurement of individual lethal doses was made; if a mouse received a dose greater than its own tolerance it died, if it received a dose less than its own tolerance it survived. The percentage kill given by any dose is therefore an estimate of the percentage of mice whose tolerances are less than that amount. Clearly the information that 16 per cent. of mice in a batch of 69 die when injected with 0·020 mg. cocaine hydrochloride

solution per gm. body-weight is less useful than a knowledge of the separate tolerances of the 69 mice, but it can nevertheless be used in a study of the distribution of individual tolerances.

80. Standardized normal distribution. Suppose that values of x, the logarithms of the individual tolerances, have a normal distribution (the same assumption that was made previously

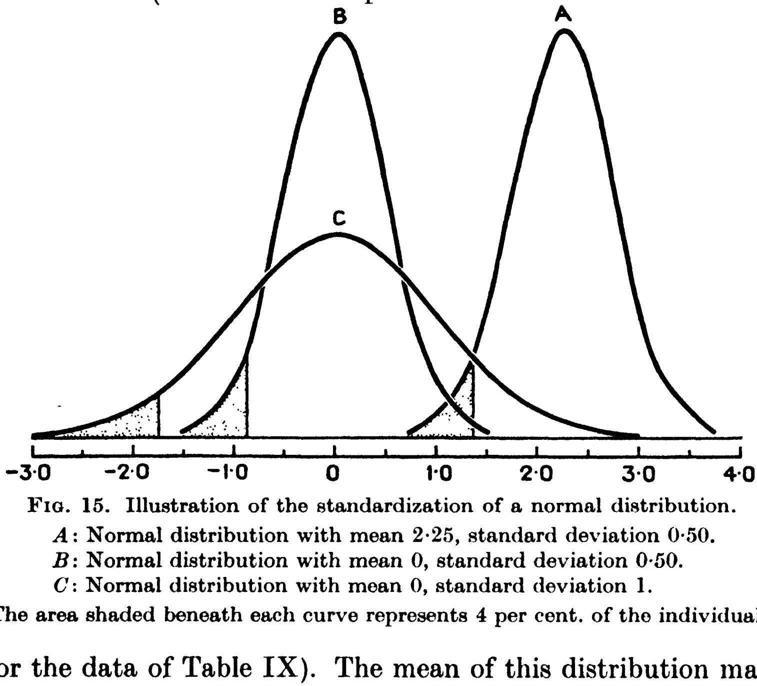


FIG. 15. Illustration of the standardization of a normal distribution.
A: Normal distribution with mean 2·25, standard deviation 0·50.
B: Normal distribution with mean 0, standard deviation 0·50.
C: Normal distribution with mean 0, standard deviation 1.
The area shaded beneath each curve represents 4 per cent. of the individuals.

for the data of Table IX). The mean of this distribution may be taken as μ and the standard deviation as σ, two quantities whose true values are unknown to the investigator but which could be estimated, if direct measurement of tolerance were possible, by $\bar{x}$ and s (§§ 8, 10). This simple estimation is not possible here, and a more detailed study of the distribution is required. If values of x have mean μ and standard deviation σ, the corresponding values of $(x-\mu)$ will have mean zero and standard deviation σ, and values of $(x-\mu)/\sigma$ will have mean zero and standard deviation 1. Furthermore, if x is normally distributed so also are $(x-\mu)$ and $(x-\mu)/\sigma$ (cf. §§ 5, 6 and Fig. 9). Associated with the normal distribution of log doses, therefore, is a *standardized normal distribution* with mean 0 and standard deviation 1. Fig. 15 illustrates the relationship between

the three distributions; curve A is a normal distribution with mean 2·25 (μ) and standard deviation 0·50 (σ), curve B is a normal distribution with mean 0 and standard deviation 0·50, and curve C is the standardized normal distribution. In each, as an example, an area representing the 4 per cent. of the distribution with the lowest values of the variate has been shaded. From curve A, the log dose required in order to kill 4 per cent. of the subjects is $x = 1{\cdot}375$; in curve B the corresponding abscissa is

$$1{\cdot}375-\mu = -0{\cdot}875,$$

and in curve C it is

$$\frac{1{\cdot}375-\mu}{\sigma} = -1{\cdot}75.$$

This last quantity is independent of μ and σ, and is determined only by the requirement that it shall correspond to 4 per cent. of the standardized normal distribution. Whatever the original distribution of tolerances, the dose which kills 4 per cent. of the subjects will correspond with an abscissa of $-1{\cdot}75$ in the standardized normal curve.

Gaddum (1933) called $-1{\cdot}75$ the *normal equivalent deviation* (N.E.D.) of 4 per cent. The equation

$$y = \frac{x-\mu}{\sigma} \tag{52}$$

is a relationship between dose and the N.E.D. of the proportion responding, which is true for any normal distribution of log tolerances, providing that the appropriate μ and σ are inserted. From experimental results such as those in Table XXV, a linear relationship between dose and the N.E.D. of the response can be estimated, and in its turn used to give estimates of μ and σ by comparison with equation (52).

81. **Probits.** The N.E.D. of any percentage less than 50 is negative. Since the publication of Gaddum's pioneer work, a slight modification in convention has been introduced in order to avoid the arithmetical inconvenience of negative numbers: instead of the N.E.D., the *probit* of the percentage is used. The probit is defined as the normal equivalent deviation increased by 5. Hereafter y is used to denote a probit, and equation (52) becomes

$$y = 5+\frac{x-\mu}{\sigma}. \tag{53}$$

The probit of any percentage greater than 0·0001 is positive. Whether or not the change is worth while is debatable, but, since all recent work has been written in terms of probits, they will be used here. Appendix VI is a simplified version of Fisher and Yates's (1948) table giving probits in terms of percentages. The relationship between Appendix VI and Table VIII should be noted. From Table VIII, 25 per cent. of the standardized normal distribution lies below $-0{\cdot}67$ and 25 per cent. above $+0{\cdot}67$; hence the probit of 25 per cent. is $(5{\cdot}00-0{\cdot}67)$, or 4·33, and the probit of $(100-25)$ per cent. is 5·67. Similarly the probit of 2·5 per cent. is $(5{\cdot}00-1{\cdot}96)$, or 3·04, and the probit of 97·5 per cent. is 6·96.

82. Probit regression line. If the percentage kill for various values of x, in the distribution represented by curve A in Fig. 15, is plotted against x, a sigmoidal relationship is found (Fig. 16): the regression of percentage kill on log dose, showing the mean kill at any dose, is a sigmoid curve, not a straight line. The experimental data plotted in Fig. 14 give some hint of this. When percentages are transformed to probits, however, the sigmoid is stretched into a straight line, in virtue of the definition of the probit. In fact, the true regression equation expressing mean kill (in probits) as a function of log dose must be

$$Y = 5+\frac{x-\mu}{\sigma}. \tag{54}$$

The straight line in Fig. 16 represents the same dose-response relationship as the sigmoid curve when the percentage scale on the left is deformed into the probit scale on the right. The calculations about to be described aim at replacing this equation, with its unknown parameters μ and σ, by an equation

$$Y = a+bx \tag{55}$$

estimated from the data as the linear regression of probit on log dose.

83. Median effective dose. Comparison of the calculated and true probit regression lines, equations (54) and (55), enables estimates of μ and σ to be formed. The mean of the tolerance distribution, μ, often described as the *median lethal dose* (LD 50) or *median effective dose* (ED 50), is the dose corresponding to a 50 per cent. response or a probit of 5. It is therefore estimated

as m, that value of x which gives $Y = 5$ in equation (55), so that

$$m = (5-a)/b. \tag{56}$$

The standard deviation of the distribution, σ, is estimated by

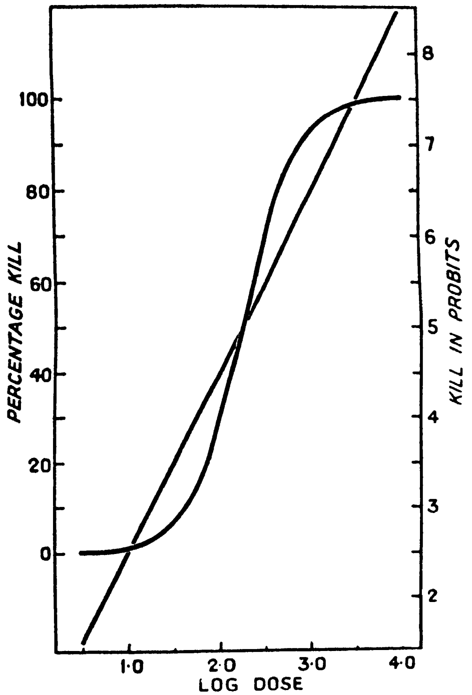


FIG. 16. The probit transformation. The sigmoid curve shows the relationship of percentage kill to log dose when distribution A of Fig. 14 is interpreted as a distribution of individual tolerances. If each percentage, on the left-hand vertical scale, is replaced by the probit, on the right-hand vertical scale (see Appendix VI), the sigmoid is transformed into the straight line shown.

the reciprocal of b, since insertion of these estimates in place of μ and σ transforms equation (54) into equation (55).

84. **Arrangement of calculations.** The calculations required for obtaining the regression line are rather more complicated than those for quantitative responses described in Section IV. Nevertheless, with the aid of Appendices VII and VIII, they can be reduced to a systematic process analogous to that of Section IV; the arithmetic is more tedious, but should have no

difficulties for those who follow the rules with care. The calculations were first set out by Bliss (1935, 1938). A detailed account of the method, with fuller descriptions of numerical examples than can be presented here, has been given by Finney (1947*a*; especially Chapters 1–5 and Appendix I).

Table XXVI shows the main calculations for fitting a probit regression line to the data in Table XXV, arranged in systematic form for routine computing. The columns have been numbered for ease of reference in the subsequent discussion, and they should be filled in the order shown. Column (1) is simply the log dose; in this example doses were multiplied by a factor of 100 (i.e. expressed as mg. per Hectogram body-weight) before the logarithms were taken, in order to avoid negative logarithms. Columns (2) and (3) contain the numbers of treated and killed animals, denoted by n and r respectively, and are copied from Table XXV. Column (4), calculated from (2) and (3), is the percentage kill, p; unless the numbers of animals at each dose are large (200 or more), p need be taken only to the nearest 1 per cent. The completion of the table is described in the paragraphs that follow.

TABLE XXVI. *Calculations for fitting a probit regression line to the data of Table XXV*

(1) x	(2) n	(3) r	(4) p	(5) Empirical probit	(6) Expected probit	(7) Weight, nw	(8) Working probit, y	(9) nwx	(10) nwy
0·176	20	0	0	..	3·0	2·6	2·58	0·4576	6·708
0·301	69	11	16	4·01	4·1	32·5	4·01	9·7825	130·325
0·398	95	50	53	5·08	5·0	60·5	5·08	24·0790	307·340
0·477	78	61	78	5·77	5·6	43·5	5·76	20·7495	250·560
0·544	44	37	84	5·99	6·2	16·3	5·97	8·8672	97·311
0·602	20	20	100	..	6·7	4·2	7·18	2·5284	30·156
Totals						159·6	..	66·4642	822·400

$\bar{x} = 0{\cdot}4164, \quad \bar{y} = 5{\cdot}1529$

$Snwx^2$	$Snwxy$	$Snwy^2$
28·8519	353·338	4,341·89
27·6785	342·482	4,237·73
1·1734	10·856	104·16
		100·44
		$3{\cdot}72 = \chi^2_{[4]}$

85. Provisional regression line. The first stage in the calculations is to enter in column (5) the probit of p, read from Appendix VI. No value can be entered for $p = 0$ or $p = 100$, since the probit there decreases or increases without limit. Fig. 17 is then prepared to show these empirical probits plotted against x. An approximately linear relationship is apparent,

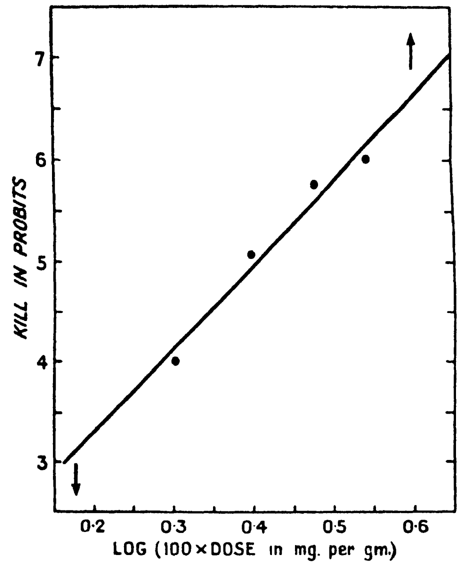


FIG. 17. Probit regression line for toxicity of cocaine hydrochloride to mice. (Data in Tables XXV and XXVI.) The arrows indicate 0 and 100 per cent. kills.

and a straight line, the *provisional regression line*, should be drawn by eye to fit the points as closely as possible. In deciding the position for this line, rather more weight should be attached to the evidence of points with probits near to 5·0 than to those with probits less than 4·0 or more than 6·0, even though all have the same value of n (§ 87). The weight attached to any point should also be proportional to n. Allowance should be made for the occurrence of a point below the line wherever $p = 0$, and a point above the line wherever $p = 100$. Skill in drawing the line increases with experience of the probit technique, and care at this stage may save labour later. A bad guess, however, cannot do any permanent harm, and at worst

will lengthen the subsequent calculations (§ 91). The line itself gives an indication of the LD 50. In Fig. 17, a probit of 5·0 corresponds to $x = 0{\cdot}405$: if the line has been well drawn, this will be approximately m, and its antilog (divided by 100 in order to return to the original units) gives 0·0254 mg. per gm. as an approximate estimate of the dose just sufficient to kill 50 per cent. Unless the points lie very close to a straight line, however, there is a danger that the provisional line may differ appreciably from the best estimate of the regression equation. This difficulty can be overcome only by completion of the remainder of Table XXVI, a step that is necessary in any case if not merely the LD 50 but also some assessment of its precision is required.

86. **Expected probits.** Column (6) contains the *expected probits*, defined to be the ordinates of the provisional line corresponding to the values of x tested; one decimal is usually sufficient. Some writers (Bliss, 1938; Fisher and Yates, 1948) recommend that the equation of the provisional regression line be found from the diagram and put in the form of equation (55), after which expected probits may be calculated by insertion of the various values of x. In practice, direct reading of the ordinates in the diagram is usually quicker, and can easily be performed to the required one-decimal accuracy. For example, in Fig. 17, the point on the provisional line for which the horizontal scale shows 0·176 is between 3·0 and 3·1, but nearer to 3·0, on the vertical scale. The expected probits are then used, in association with Appendices VII and VIII, to give the *weights* (§ 87) and *working probits* (§ 88). No attempt is made here to provide a full account of these quantities; Finney (1947*a*) may be consulted for the theory of their derivation and for statements of their algebraic formulae.

87. **Weights of observations.** Suppose that two percentage responses, say, 5 per cent. and 45 per cent., are each determined from the same number of individuals. The variance of the first on account of the binomial distribution will be less than that of the second by virtue of equation (47). Nevertheless, because of the slope of the sigmoid percentage response curve, illustrated in Fig. 16, the range of doses which may produce a 5 per cent. response is wider than that of doses which may produce a 45 per cent. response. Hence an observed response of 5 per cent.

gives less information about the dose which produced it than does an observed response of 45 per cent. In probit calculations, allowance must be made for this by allocating a *weight* to each response, and so to each probit, according to the value of the expected probit. Appendix VII contains the *weighting coefficient*, w, for different values of the expected probit. It is the weight per individual subjected to treatment, and must be multiplied by n in order to give the total weight of an observation for entry in column (7). The coefficient w has a maximum at a probit of 5·0 (i.e. 50 per cent. response) and falls off rapidly outside the range 4·0 to 6·0. At an expected probit of 4·1, for example, $w = 0{\cdot}471$, and therefore the weight for the second dose in Table XXVI is

$$nw = 69 \times 0{\cdot}471 = 32{\cdot}5.$$

The total weight, Snw, is entered at the bottom of the column.

88. **Working probits.** The working probits in column (8) enable the right use to be made of the information provided by the extreme doses that kill 0 per cent. or 100 per cent.; they also take account of the non-linearity of the transformation of all the percentage death-rates to their probits. Working probits, rather than empirical probits, are denoted by y, since they are the values used in the regression calculations, and the formulae then have a close analogy with those of Section IV. Each value of y is obtained by entering Appendix VIII in the line for the appropriate value of the expected probit and in the column corresponding to the percentage kill, p.

The table in Appendix VIII is a condensed version of one given by Finney (1947*a*, Table IV), and shows p only at intervals of 5 per cent. Values can be interpolated rapidly, and almost by inspection after some experience of the table. For example, the working probit for the first dose is found in the line of Appendix VIII for an expected probit 3·0 and in the column for $p = 0$. The last dose has an expected probit 6·7 and $p = 100$, a combination which is not explicitly shown in Appendix VIII; however, for an expected probit 6·6, $y = 7{\cdot}09$, and for an expected probit 6·8, $y = 7{\cdot}26$, whence the required value is

$$y = \tfrac{1}{2} \times (7{\cdot}09 + 7{\cdot}26)$$
$$= 7{\cdot}18.$$

The fifth dose has an expected probit 6·2 and $p = 84$. Now

$$\text{for } p = 80,\ y = 5{\cdot}76,$$
$$\text{and for } p = 85,\ y = 6{\cdot}02;$$

hence the required y is

$$\begin{aligned} y &= 5{\cdot}76+\tfrac{4}{5}\times(6{\cdot}02-5{\cdot}76)\\ &= 5{\cdot}76+\tfrac{4}{5}\times 0{\cdot}26\\ &= 5{\cdot}97. \end{aligned}$$

An alternative form of calculation, dependent upon the fact that 84 divides the interval from 80 to 85 in the ratio 8:2, is more convenient with a calculating machine:

$$\begin{aligned} y &= 0{\cdot}2\times 5{\cdot}76+0{\cdot}8\times 6{\cdot}02\\ &= 5{\cdot}97. \end{aligned}$$

The most troublesome form of interpolation is that for the second dose, with an expected probit of 4·1 and $p = 16$. The procedure just described gives

for 4·0 $\quad y = 0{\cdot}8\times 3{\cdot}96+0{\cdot}2\times 4{\cdot}17 = 4{\cdot}00$

and for 4·2 $\quad y = 0{\cdot}8\times 3{\cdot}99+0{\cdot}2\times 4{\cdot}16 = 4{\cdot}02$

so that for an expected probit of 4·1

$$\begin{aligned} y &= \tfrac{1}{2}\times(4{\cdot}00+4{\cdot}02)\\ &= 4{\cdot}01. \end{aligned}$$

A more convenient calculation is

$$\begin{aligned} y &= 0{\cdot}4\times(3{\cdot}96+3{\cdot}99)+0{\cdot}1\times(4{\cdot}17+4{\cdot}16)\\ &= 4{\cdot}01. \end{aligned}$$

The interpolations take longer to describe than to do, and the method should be clear to any who will work through these three examples. Values of y at intervals of 0·1 in the empirical probit and 1 in p can be read directly from Finney's table.

It will be noted from inspection of Table XXVI that columns (5) and (8) are almost identical. Indeed, in this example, no harm would have resulted from neglect of working probits (except for the two extreme doses) and use of the empirical probits instead; this will always be so when the empirical and

expected probits are close together. If a point in Fig. 17 lay far from the line, so that the empirical and expected probits differed widely, the working probit would be very different from the empirical, but otherwise a large difference should be re-examined under the suspicion that it is due to an arithmetical fault. Though empirical probits might sometimes be used directly with impunity, the general rule of always using working probits is to be recommended. This correct procedure is particularly important when small values of n permit the occurrence of very erratic percentage kills.

Some writers use the term 'corrected probit' instead of working probit, but this is an unfortunate choice of name. The working probit is used as an aid to the performance of otherwise complicated calculations, and is in no sense a correction of the empirical value. Formulae for potency estimates and their variances are calculated from working probits; as may be proved by more advanced statistical theory, they are none the less the genuine deductions from the data and do not imply a 'cooking' of the observations in order to remove inconvenient irregularities!

89. Regression calculations. In Fig. 17, as an essential step to the calculations made so far, a regression line was drawn by eye. Calculation of a *weighted regression equation* of y on x can be shown to give an improved version of the line. By a weighted regression equation is meant one in which, instead of every x and y being of equal importance in the calculations, each value is used in association with the corresponding weight, nw. The role of $S(x-\bar{x})^2$ in the unweighted regression calculations (§ 23) is now played by $Snw(x-\bar{x})^2$. Similarly $S(x-\bar{x})(y-\bar{y})$ is replaced by $Snw(x-\bar{x})(y-\bar{y})$ and $S(y-\bar{y})^2$ by $Snw(y-\bar{y})^2$. In the new expressions, $\bar{x}$ and $\bar{y}$ are to be taken as weighted means.

The arithmetic is a little more troublesome than that described in Section IV. As a start, columns (9) and (10) of Table XXVI must be completed by entering in them the products nwx and nwy. For example, for the first dose

$$2{\cdot}6 \times 0{\cdot}176 = 0{\cdot}4576,$$

$$2{\cdot}6 \times 2{\cdot}58 = 6{\cdot}708.$$

The sums of the columns are $Snwx$, $Snwy$, from which the

weighted means are derived by formulae analogous to equation (1):

$$\bar{x} = \frac{Snwx}{Snw} \tag{57}$$
$$= 66{\cdot}4642/159{\cdot}6$$
$$= 0{\cdot}4164,$$

and

$$\bar{y} = \frac{Snwy}{Snw}$$
$$= 822{\cdot}400/159{\cdot}6$$
$$= 5{\cdot}1529.$$

The sums of squares and products, which may again be symbolized by S_{xx}, S_{xy}, S_{yy}, could be found by tabulation of the various differences $(x-\bar{x})$ and $(y-\bar{y})$, followed by direct multiplication and summation. Generalizations of equations (5) and (17), however, lead to less laborious calculations. In fact,

$$S_{xx} = Snw(x-\bar{x})^2 = Snwx^2 - \frac{(Snwx)^2}{Snw} \tag{58}$$
$$= 28{\cdot}8519 - \frac{(66{\cdot}4642)^2}{159{\cdot}6}$$
$$= 1{\cdot}1734.$$

The expression $Snwx^2$ is obtained by multiplying each entry in column (9) by the corresponding entry in column (1) and summing the six products. A sum of products of entries in columns (9) and (8) gives $Snwxy$, and the result should be checked by the alternative calculation as the sum of products of columns (10) and (1); it is then used in evaluating

$$S_{xy} = Snw(x-\bar{x})(y-\bar{y}) = Snwxy - \frac{(Snwx)(Snwy)}{Snw} \tag{59}$$
$$= 353{\cdot}338 - \frac{66{\cdot}4642 \times 822{\cdot}400}{159{\cdot}6}$$
$$= 10{\cdot}856.$$

The sum of products of columns (10) and (8), $Snwy^2$, is used as was $Snwx^2$, and gives

$$S_{yy} = Snw(y-\bar{y})^2 = Snwy^2 - \frac{(Snwy)^2}{Snw}$$
$$= 4{,}341{\cdot}89 - \frac{(822{\cdot}400)^2}{159{\cdot}6}$$
$$= 104{\cdot}16.$$

All these calculations are summarized in Table XXVI, in the position in which they would normally appear in routine computing.

90. Regression equation. The modified definitions of S_{xx}, S_{xy}, and S_{yy} enable the regression coefficient of y on x to be expressed by the familiar formula (14):

$$\begin{aligned} b &= S_{xy}/S_{xx} \\ &= 10{\cdot}856/1{\cdot}1734 \\ &= 9{\cdot}252. \end{aligned}$$

The regression equation is again equation (13):

$$\begin{aligned} Y &= \bar{y}+b(x-\bar{x}), \\ \text{or } Y &= 5{\cdot}1529+9{\cdot}252(x-0{\cdot}4164), \\ \text{or } Y &= 1{\cdot}300+9{\cdot}252x. \end{aligned}$$

If the six values of x (Table XXVI) are inserted in turn in this equation, the corresponding probits are found to be 2·93, 4·08, 4·98, 5·71, 6·33, and 6·87; the six points lie on a straight line, the calculated regression line, which, on the scale of Fig. 17, would scarcely be distinguishable from the provisional line shown there.

91. Iterative calculations. Had the regression line whose equation was found in § 90 been sensibly different from the provisional line, further calculation on an iterative scheme would have been needed. The six values of Y calculated at the end of § 90 (or these reduced to 1 decimal) would have been inserted as a new series of expected probits in a revised version of Table XXVI. The calculations described in §§ 87–90 would then have been repeated in order to give a new regression line. This line would give another series of expected probits, and additional cycles of calculation would be completed until close agreement between two successive cycles was obtained. The last regression line would be that used to represent the dose-response relationship. In practice, a difference of 0·2 or less between an entry in column (6) and the new value obtained from the subsequently calculated line is unimportant unless the data are very extensive (i.e. n is large). Care in the drawing of the provisional line should ensure that at most two cycles of

calculations are needed; often one suffices, as in the numerical example now under discussion.

92.† Sigmoid response curve. In general, construction of the curve relating percentage response to dose is unnecessary, but it serves here to illustrate the meaning of the regression line, and to show the close agreement between the data and the law of variation calculated for them. Values of Y can be calculated from the regression equation for a series of values of x, not necessarily only those used in the experiment. Appendix VI is used to give the percentages corresponding to these probits, which are then plotted against x. The result is the sigmoid curve in Fig. 14, the estimated regression relationship of percentage kill on log dose.

93. Test of linearity. The next important step is to apply a test of validity to the hypothesis on which the whole probit method rests—that the individual tolerances, on the log dose scale, are normally distributed. This takes the form of a χ^2 test, but is closely related to the analysis of variance test for examining the linearity of a regression, described in § 26. Formula (19) again gives a component of the total (weighted) sum of squares for y, with 1 degree of freedom, attributable to the regression. Hence, if the experiment involves k levels of dose,

$$S_{yy}-\frac{(S_{xy})^2}{S_{xx}} \tag{60}$$

is a residual sum of squares with $(k-2)$ degrees of freedom, dependent upon deviations from the probit regression line. Here, apparently, there is no error mean square with which to compare a mean square derived from formula (60). The theory of the probit calculations, however, depends in part upon the binomial distribution, for which a true, or theoretical, variance (based upon expected frequencies) is given by equation (47). A consequence, admittedly by no means obvious without a full algebraic argument, is that, if the individual tolerances are distributed normally and independently of one another, formula (60) represents a χ^2, and its significance can be tested by means of Appendix V. Here

$$\chi^2_{[4]} = 104{\cdot}16-\frac{(10{\cdot}856)^2}{1{\cdot}1734}$$
$$= 3{\cdot}72,$$

the subscript 4 indicating the degrees of freedom. A value as great as 9·5 could occur by chance (Appendix V; 0·05 probability), and thus there is no suspicion of non-normality, or, what is here another aspect of the same thing, non-linearity of the probit regression line.

This procedure is strictly correct only after the limit of the iterative calculations described in § 91 has been attained, but, provided that the latest cycle has given a regression line almost the same as the provisional line used in that cycle, little inaccuracy need be feared. Formula (60) can be shown to be the same, in the limit, as $Snw(y-Y)^2$, the weighted sum of squares of deviations from the regression line (cf. § 25). Further algebraic rearrangement, with the aid of the formulae for weighting coefficients and working probits, enables this to be re-written

$$S\left\{\frac{(np-nP)^2}{nP}+\frac{(nq-nQ)^2}{nQ}\right\}; \tag{61}$$

P, Q are expected probabilities of death and survival respectively, derived from the values of Y for each dose by means of Appendix IV. For example, from § 90 the second dose has $Y = 4{\cdot}08$, which corresponds to $P = 0{\cdot}18$ and $Q = 0{\cdot}82$. Formula (61) is the same as the formula for χ^2 in a contingency table, equation (51). Herein lies the justification for testing the expression (60) as a χ^2. An additional degree of freedom is lost because not merely an average death-rate but also a regression on dose has been estimated from the data.

The test is likely to be upset by the occurrence of small expected frequencies: a 'significant' χ^2 may arise because one or two classes have very small expected frequencies, and not because the probit analysis is invalid. If there is any reason to suspect that a large χ^2 has arisen in this way, a more detailed comparison of observed and expected frequencies should be made; an account of the method has been given elsewhere (Finney, 1947*a*, § 18).

94. Heterogeneity. A warning must be given against any temptation to ignore the χ^2 test. A large and significant χ^2 may denote non-linearity of the probit-dose relationship, in which case the method of analysis used here is inappropriate, and expert advice on a substitute should be sought. Alternatively, it may arise because of 'heterogeneity between batches':

the animals tested at any one dose may be less variable than those tested at different doses, because of some lack of completely random selection for testing (§ 132). The two cases can usually be distinguished by inspection of a diagram such as Fig. 17, with the calculated regression line drawn on it. The first would show itself as a systematic departure from the line indicating a regular curvature, the second would still allow the line as the best fit to the points, but the points would be scattered more widely than the theory of the binomial distribution predicts, so leading to the large χ^2. In the latter circumstances, a *heterogeneity factor* defined as χ^2 divided by its number of degrees of freedom, $(k-2)$, must be introduced as a multiplier of g and of all expressions for variances, and a t-value for $(k-2)$ degrees of freedom must replace 1·96; this modification applies to all the calculations described in § 96 (Finney, 1947*a*, §§ 11, 18).

95. Estimation of median lethal dose. As the χ^2 test has shown no reason for doubting the validity of the probit technique in the analysis of the cocaine hydrochloride toxicity data, estimation may proceed. The LD 50 is given by equation (56), which may be written

$$\begin{aligned} m &= \bar{x}+\frac{5-\bar{y}}{b} \qquad (62) \\ &= 0{\cdot}4164-\frac{0{\cdot}1529}{9{\cdot}252} \\ &= 0{\cdot}3999; \end{aligned}$$

the result is very close to 0·405, the value obtained in § 85 from the provisional line. Its antilogarithm gives 0·0251 mg. per gm. as the dose estimated to kill 50 per cent. of mice.

96. Precision of estimated LD 50. Equation (62) shows that m involves a ratio of two estimates from the data, $\bar{y}$ and b. Unless b is considerably greater than its standard error, its occurrence in the denominator of m will prevent the use of a simple formula for the standard error of m; the difficulty is the same as that encountered in § 18 and § 39 for R and M respectively.

In § 93 it was explained that the test of linearity of regression, usually a variance ratio test, becomes a χ^2 test in probit calculations because the binomial distribution provides a theoretical error variance. A second effect of the binomial

distribution, in combination with the weighting coefficients used, is to replace the variance estimate s^2 (Sections IV–VI) by a theoretical variance of unity. Equation (21) thus becomes

$$V(b) = \frac{1}{S_{xx}} \tag{63}$$
$$= 1/1{\cdot}1734$$
$$= 0{\cdot}852,$$

so that b ($= 9{\cdot}25$; § 90) is ten times its standard error. The g-criterion may be used as in § 39 and § 52, except that the theoretical variance must be inserted in equation (26), and consequently a normal deviate (or a t with an unlimited number of degrees of freedom) must be used for t:

$$g = \frac{t^2}{b^2 S_{xx}} \tag{64}$$
$$= \frac{(1{\cdot}96)^2}{(9{\cdot}25)^2 \times 1{\cdot}173}$$
$$= 0{\cdot}038$$

at the 0·05 probability level. Since g is less than 0·1, an approximate variance of m may be used. The formula is analogous to equation (33), with s^2 replaced by unity; it is

$$V(m) = \frac{1}{b^2}\left[\frac{1}{Snw} + \frac{(m-\bar{x})^2}{S_{xx}}\right] \tag{65}$$
$$= \frac{1}{(9{\cdot}252)^2} \times \left[\frac{1}{159{\cdot}6} + \frac{(0{\cdot}0165)^2}{1{\cdot}173}\right]$$
$$= 0{\cdot}00007575.$$

The square root of this gives the standard error of m, and therefore

$$m = 0{\cdot}3999 \pm 0{\cdot}0087.$$

The establishment of fiducial limits to the LD 50 involves, implicitly, the use of a normal distribution as an approximation to a binomial. As in § 73, the normal deviate must then be used, and not a t-deviate.[1] The limits will therefore be placed at

$$1{\cdot}96 \times 0{\cdot}0087 = 0{\cdot}0171$$

[1] A t-deviate with $(k-2)$ degrees of freedom would be quite inappropriate, for the binomial variance is based upon variation within doses, not between doses. When the data show heterogeneity, however, much of this paragraph must be modified, as stated in § 94.

on either side of m, that is to say, at 0·3828 and 0·4170. The antilogarithms of these are 2·41 and 2·61. The median lethal dose is *estimated as 0·0251 mg. per gm.* body-weight, and may be *confidently asserted to be greater than* 0·0241 mg. per gm., *less than* 0·0261 mg. per gm.

Had g not been small, a more complicated expression for the fiducial limits would have been needed. This is very similar to that used in § 52; it will not be quoted here as, in fact, limits of m itself are seldom required, and an extension of the result for assay purposes will be discussed in § 107.

97.† Natural mortality. Dose-mortality relationships are sometimes affected by the occurrence of a natural mortality amongst the test subjects independent of that due to the dosage. If the quantal response studied is not death, a similar difficulty —the occurrence of responses observationally indistinguishable from those studied, but actually independent of the dose—may arise. Unless this natural response rate is very small, the statistical analysis must be altered to take account of it. Description of the requisite *adjustment for natural mortality* cannot be given here, and the reader must be referred to other publications (Finney, 1947*a*, Chapter 6; Fisher and Yates, 1948, Table XI_1).

98. Note on computing. The calculations for a probit regression line may appear excessively complicated at a first reading. Systematic arrangement and a little experience will remove the terrors, and the whole of the calculations just described will then be completed easily and rapidly. Finney (1947*a*, Appendix I) has described in detail a systematic scheme of computation with checks at every step, and this or some equivalent scheme should be adopted as a standard by those who have to perform like analyses frequently. The reader to whom the probit method is new is strongly urged to repeat for himself every step of the calculations in this section, following the instructions with care, and to make similar repetitions of the main calculations in Section X, before deciding that the method is too difficult for him!

X. Standardization using Quantal Responses

99. Principles of standardization. The probit may be regarded as a response metameter (§ 67) which, when used with percentage responses, converts a sigmoidal regression on log dose

into a linear regression. The argument developed in Sections V and VI shows again that if a standard and a test preparation produce their effects by means of the same active principle, the two regression lines must be parallel, and the horizontal distance between them estimates the logarithm of their relative potency. The statistical methods discussed in these sections are therefore applicable also to assays based upon quantal responses, except for some modification in the form of the calculations on account of the use of weighted regressions. The principles of the analysis are unaltered and will not be discussed again in detail; the only new difficulties are those of the probit transformation, and the chief features of this have already been explained in Section IX.

100. Standard curve. In theory an assay procedure could be based upon reference to a response curve determined for the standard, in the manner described in § 37. In practice, the same objection as before arises: the assumption that the standard curve remains unchanged from day to day or from laboratory to laboratory is manifestly false for many dose-response relationships.

101.† Characteristic curve. Corresponding to the standard regression coefficient method described in § 38 is a method using what Trevan (1927) has termed the *characteristic curve*. A sigmoid curve relating percentage response to dose of a standard preparation is constructed, either by freehand drawing in a diagram such as Fig. 14 or by probit calculations, and from it are read the doses corresponding to various response percentages. These are then expressed as percentages of the LD 50. The resulting relationship between dose as a percentage of LD 50 and percentage of subjects responding is regarded as the characteristic curve for use in future standardization. When a test preparation is to be assayed, a single dose of it and the standard can be tested simultaneously, and the responses in the two batches referred to the characteristic curve in order to give potency percentages; the ratio of these two then gives the required relative potency.

The method is in all essential features the same as the standard regression method, and an example will make clear its operation. Table XXVII shows doses of cocaine hydrochloride read from the sigmoid curve in Fig. 14 at intervals of 5 per cent. in response, and these same doses expressed as percentages of the

LD 50. A plotting of the one set of percentages against the other would give the characteristic curve, but there is no necessity actually to draw this. Suppose, now, that a dose having 0·1 mg. of a test preparation containing cocaine per gm. body-weight kills 15 mice in a batch of 20, and that a simultaneous trial of 0·02 mg. standard solution of cocaine hydrochloride per gm. kills 7 mice out of 15. The test preparation produced a mortality of 75 per cent. and the standard preparation a mortality of 47 per cent. From the characteristic curve, or from Table XXVII, the potencies corresponding to these are 118 and 98 respectively. Hence 0·1 mg. of the test preparation is estimated to contain the equivalent of

$$0{\cdot}02 \times \frac{118}{98} \text{ mg. cocaine hydrochloride solution}$$

or 1 mg. of the test preparation is equivalent to 0·24 mg. cocaine hydrochloride solution.

TABLE XXVII. *Specification of characteristic curve for cocaine hydrochloride, from data of Table XXV*

Per cent. mortality	*Dose (mg. per gm.)*	*Per cent. potency*	*Per cent. mortality*	*Dose (mg. per gm.)*	*Per cent. potency*
5	0·0167	67	55	0·0259	103
10	0·0183	73	60	0·0267	106
15	0·0194	77	65	0·0277	110
20	0·0204	81	70	0·0286	114
25	0·0212	85	75	0·0296	118
30	0·0221	88	80	0·0310	123
35	0·0228	91	85	0·0325	129
40	0·0236	94	90	0·0345	137
45	0·0243	97	95	0·0378	150
50	0·0251	100			

The validity of the method depends upon the constancy of the characteristic curve, which can be shown to be the same thing as the constancy of the variance of individual log tolerances. Different stocks of animals, bred under different conditions, will not necessarily have the same tolerance variances, so that a characteristic curve determined in one laboratory may not be applicable in another. Even within one laboratory, there can be little assurance that the variance will remain unaltered for an indefinite period. The method of assay provides no

evidence of a change in the characteristic curve, and seriously misleading conclusions might be drawn from use of an inappropriate curve; the danger is greatest when the percentage responses for standard and test preparation differ widely. As stated in § 44, a self-contained assay procedure is to be preferred unless the investigator is certain of the applicability of his characteristic curve. For this reason, the standard error of a relative potency estimate based upon a characteristic curve will not be discussed here, though all that is required is a slight modification of the procedure for the standard regression coefficient method described in § 39.

If the characteristic curve method were found to be valid for some type of assay, the initial determination of the curve ought to be made with considerable care. A curve that is to be used for many assays should be determined fairly exactly, from more doses and larger numbers of subjects than the cocaine hydrochloride data. In theory, the application of the method need not then be restricted to the approximately linear part of the curve, though the consequences of slight variations in the variance of the tolerance distribution would be the most serious in assays involving very high or very low responses.

102. A 4-point assay. Gaddum (1933) reported the data in Table XXVIII as the results of two assays of strophanthus by use of the percentage mortality observed in frogs. In the first assay (Table XXVIII*a*), two doses of the standard and test

TABLE XXVIII. *Mortality of frogs* (Rana temporaria) *after injection with strophanthus*

(*a*) *Batches of 15 frogs*

	Concentration of tincture, per litre saline			
	6 c.c. S	*7 c.c. S*	*6 c.c. T*	*7 c.c. T*
No. killed	11	12	3	6

(*b*) *Batches of 20 frogs*

	Concentration of tincture, per litre saline			
	4 c.c. S	*6 c.c. S*	*6 c.c. T*	*9 c.c. T*
No. killed	2	16	7	19

tinctures, S and T, were used, each being injected subcutaneously into frogs at a rate of 0·02 c.c. per gm. body-weight (cf. § 31). The results show no significant evidence of an increase in mortality with increasing dose, and are therefore useless for standardization purposes. This may be shown by a test of significance of the probit regression coefficient, after calculations similar to those about to be described for the second assay. More easily, a rough test may be based upon a pooling of the S and T results into a 2×2 contingency table:

	Alive	*Dead*	*Total*
Low dose	16	14	30
High dose	12	18	30
Total	28	32	60

Calculation by the method of § 76 gives

$$\chi^2 = (1{\cdot}5)^2 \times (\tfrac{2}{14}+\tfrac{2}{16})$$
$$= 0{\cdot}6.$$

The χ^2 test is open to objection because of the combination of the S and T series, but here its indications are perfectly clear: there is no significant evidence that the mortality from the high dose is greater than that from the low dose.

Though the first assay cannot itself lead to an estimate of potency, it has served a useful purpose by indicating that the dose interval used was too narrow and that the doses of S were too high. These faults were corrected in the second assay, in which the interval between high and low log doses was 3 times as great as before. Again the tinctures were injected at a rate of 0·02 c.c. per gm. body-weight, and 20 frogs were used in each of the 4 groups. The statistical analysis of the results of the second assay will now be described.

103. Regression calculations. The calculations required for applying the probit method to the estimation of relative potency are shown in Table XXIX. The arrangement is similar to that of Table XXVI. Columns (1), (2), and (3) contain respectively the log dose (x), the number of subjects at that dose (n), and the number killed (r), these being taken from Table XXVIII. A change of the x scale could be introduced here, in order to

simplify later calculations, and a compensating adjustment made at the end: the interval between log doses is the same for both preparations, and the values of x might therefore be taken simply as 0,1 for each preparation. In order to show the calculations in their more general form, as required when the intervals in log doses are unequal, this procedure has not been adopted.

TABLE XXIX. *Calculations for fitting parallel probit regression lines to the data of Table XXVIII (b)*

(1) x	(2) n	(3) r	(4) p	(5) *Empirical probit*	(6) *Expected probit*	(7) *Weight, nw*	(8) *Working probit, y*	(9) nwx	(10) nwy
Standard preparation									
0·602	20	2	10	3·72	3·8	7·4	3·72	4·4548	27·528
0·778	20	16	80	5·84	5·8	10·1	5·84	7·8578	58·984
Totals						17·5	..	12·3126	86·512
Test preparation									
0·778	20	7	35	4·61	4·6	12·0	4·61	9·3360	55·320
0·954	20	19	95	6·64	6·6	4·8	6·64	4·5792	31·872
Totals						16·8	..	13·9152	87·192

$$\bar{x}_S = 0{\cdot}7036, \quad \bar{y}_S = 4{\cdot}944,$$
$$\bar{x}_T = 0{\cdot}8283, \quad \bar{y}_T = 5{\cdot}190.$$

	$Snwx^2$	$Snwxy$	$Snwy^2$
Standard:	8·7952	62·461	446·87
	8·6629	60·868	427·68
	0·1323	1·593	19·19
Test:	11·6320	73·445	466·66
	11·5258	72·220	452·53
	0·1062	1·225	14·13
Totals:	0·2385	2·818	33·32
			33·30
			$0{\cdot}02 = \chi^2_{[1]}$

Column (4) shows the percentage kill, and this is converted to an empirical probit, by means of Appendix VI, for entry in column (5). Empirical probits are plotted against x in Fig. 18. The two pairs of points in this diagram suggest the possibility of parallel regression lines for the two preparations, and, with this in view, provisional lines are drawn parallel in positions judged by eye to fit the data reasonably well; again experience

in the probit technique aids this judgement, though the only ill effect a bad choice can have is to lengthen the calculations (§ 91). The horizontal distance between these lines gives a first estimate of M, just as in § 47:

$$M = -0{\cdot}107,$$

a result very close to that obtained in § 106 after the full

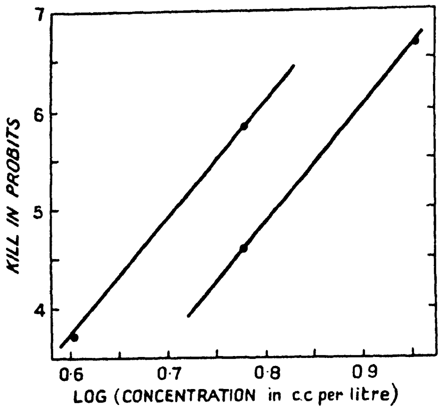


FIG. 18. Assay of strophanthus from mortality of frogs. (Data in Tables XXVIII (*b*) and XXIX.)
Upper line: Standard preparation.
Lower line: Test preparation.

regression calculations. The calculations are needed, however, if the precision of the estimate of potency is to be assessed.[1]

The points on the provisional lines at the values of x tested give the expected probits (§ 86) for column (6). Since n is the same for all the doses in the assay, the calculations might be carried out with the weighting coefficient, w, used in place of nw; the multiplier, n, would then be introduced only when the sums of squares and products had been formed. Data in which n is not the same for all doses occur so frequently, however, that for illustrative purposes the more general procedure seems

[1] It is possible to use the diagram, in conjunction with certain standard nomographs, to simplify some stages of the regression calculations, though the method is not exactly correct. An understanding of the orthodox procedure is desirable before such a modification is adopted. The method, due to de Beer (1945), will not be described here, but may be useful for routine assays, providing that the data are first carefully inspected in order to detect any anomalies.

preferable. From Appendix VII, the weighting coefficient at an expected probit of 3·8 is 0·370, and therefore

$$nw = 20 \times 0{\cdot}370 = 7{\cdot}4$$

for the lower dose of S. From Appendix VIII, in the row for 3·8 and the column for 10 per cent., the working probit is read as 3·72. So columns (7) and (8) are completed, in the manner described more fully in §§ 87, 88.

The products nwx and nwy are entered in columns (9) and (10), and totalled separately for each preparation. From them are obtained means $\bar{x}$ and $\bar{y}$, by equation (57):

$$\bar{x}_S = \frac{12{\cdot}3126}{17{\cdot}5}$$
$$= 0{\cdot}7036, \text{ etc.}$$

Sums of squares and products are obtained just as in § 89, and recorded at the bottom of the table. For example, by equation (59) for the standard preparation,

$$\begin{aligned} S_{xy} &= 3{\cdot}72 \times 4{\cdot}4548 + 5{\cdot}84 \times 7{\cdot}8578 - \frac{12{\cdot}3126 \times 86{\cdot}512}{17{\cdot}5} \\ &= 62{\cdot}461 - 60{\cdot}868 \\ &= 1{\cdot}593, \end{aligned}$$

and a check may be applied by forming the sum of products of nwy with x.

104. Validity test. The validity test required here is closely analogous to that for the 4-point assay with quantitative responses discussed in Section VI, but a χ^2 test is used, as in § 93, instead of the variance ratio test in § 49. As always with a 4-point design, the data lead to a test of parallelism of the regression lines but cannot provide a test of linearity. (More strictly, the linearity and parallelism tests are not separable, for in a 4-point assay non-linearity might manifest itself as non-parallelism if the levels of response for the two preparations were different.) In fact, each S_{yy} is based on only 1 degree of freedom, and an attempt to form a χ^2 for testing linearity, by formula (60), must give zero, since there are no degrees of freedom for the χ^2. The relation

$$S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 0$$

may be used as an additional check on the calculations of Table XXIX. If the same procedure is followed on the pooled sums of squares and products at the bottom of Table XXIX, however, since these have 2 degrees of freedom, the resulting χ^2 has 1 degree of freedom:

$$\chi^2_{[1]} = 33{\cdot}32 - \frac{(2{\cdot}818)^2}{0{\cdot}2385}$$
$$= 0{\cdot}02;$$

the χ^2 tests the significance of the deviation from parallelism of regression lines fitted to the two preparations separately. In this example, there is no indication of non-parallelism to suggest that the assay may be invalid.

Though logically the validity test needs to be considered before the regression equations are used, the test cannot properly be made until the iterative calculations (§ 91) have been carried as far as is necessary in order to give reasonably stable positions for the regression lines. The calculation of b and the regression equations (§ 105) is therefore usually completed before the validity test.

105. Regression equations. The average regression coefficient is obtained (cf. §§ 48, 58) by pooling the values of S_{xx} and of S_{xy} to give

$$b = \frac{1{\cdot}593 + 1{\cdot}225}{0{\cdot}1323 + 0{\cdot}1062}$$
$$= \frac{2{\cdot}818}{0{\cdot}2385}$$
$$= 11{\cdot}82.$$

By equation (13), the revised regression equations are

$$Y_S = 4{\cdot}944 + 11{\cdot}82(x - 0{\cdot}7036)$$
$$= -3{\cdot}37 + 11{\cdot}82x,$$

and $$Y_T = -4{\cdot}60 + 11{\cdot}82x.$$

If the values of x in Table XXIX are substituted in the appropriate equation, probits for the fitted lines are found to be 3·75, 5·83 for S and 4·60, 6·68 for T, values so close to the expected probits in column (6) of Table XXIX that clearly no second cycle of calculations is needed. Had there been substantial differences, as there might have been if the provisional lines in Fig. 18 had been badly judged, the whole analysis would

have been recomputed, with a new column (6) consisting of the probits just calculated, in the manner described in § 91.

106. Estimation of relative potency. The logarithm of the relative potency is, as usual, the difference between the logarithms of equally effective doses of the two preparations. If this difference is to be interpreted in terms of relative potency, it must be independent of the level of response at which the comparison is made: as has already been stated, the two regression lines must be parallel, and the logarithm of the relative potency is then estimated by the horizontal distance between them. This logarithm is, therefore, still given by equation (32):

$$M = \bar{x}_S - \bar{x}_T - \frac{\bar{y}_S - \bar{y}_T}{b},$$

though the $\bar{x}$ and $\bar{y}$ must now be understood as weighted means. For the strophanthus assay

$$\begin{aligned} M &= 0{\cdot}7036 - 0{\cdot}8283 - \frac{4{\cdot}944 - 5{\cdot}190}{11{\cdot}82} \\ &= -0{\cdot}1247 + 0{\cdot}0208 \\ &= -0{\cdot}1039 \\ &= \bar{1}{\cdot}8961, \end{aligned}$$

the antilogarithm of which gives

$$R = 0{\cdot}787.$$

Hence 1 c.c. of the test tincture of strophanthus is estimated to be equivalent to 0·787 c.c. of the standard tincture.

107. Precision of potency estimate. The formula for M is similar to that for a log LD 50 (§ 95), and also involves b in its denominator. Once again, therefore, special attention must be paid to the standard error of b; more conveniently, in practice, a g-criterion is usually applied in order to decide whether or not an approximate formula for $V(M)$ may be used. The variance of b is given by equation (63), in which S_{xx} must be understood as the pooled sum of squares for log dose at the bottom of Table XXIX:

$$\begin{aligned} V(b) &= \frac{1}{0{\cdot}2385} \\ &= (2{\cdot}05)^2, \end{aligned}$$

so that b is between five and six times its standard error.

The approximate variance of M is given by equation (33), except that N, the number of subjects for either preparation, must be replaced by Snw, the total weight of the observations for that preparation, and s^2 must be replaced by unity (§ 96). Hence

$$V(M) = \frac{1}{b^2}\left\{\frac{1}{{}_SSnw}+\frac{1}{{}_TSnw}+\frac{(M-\bar{x}_S+\bar{x}_T)^2}{S_{xx}}\right\}, \qquad (66)$$

where ${}_SSnw$ and ${}_TSnw$ refer to the preparations and S_{xx} is the pooled sum of squares. For many assays, especially when large numbers of subjects have been used, so that b is well determined, fiducial limits can be derived from equation (66). Before this is done, however, the g-criterion should be applied; it is again given by equation (64):

$$\begin{aligned} g &= \frac{t^2}{b^2 S_{xx}} \\ &= \frac{(1{\cdot}96)^2}{(11{\cdot}82)^2 \times 0{\cdot}2385} \\ &= 0{\cdot}115. \end{aligned}$$

Clearly g is too large for $V(M)$ to be safely used. The exact formula required for the lower fiducial limit to M, derived from formula (35) by the same rule as was used above for $V(M)$, is

$$\begin{aligned} &\bar{x}_S-\bar{x}_T+\frac{M-\bar{x}_S+\bar{x}_T}{1-g} \\ &\qquad -\frac{t}{b(1-g)}\sqrt{\left\{(1-g)\left(\frac{1}{{}_SSnw}+\frac{1}{{}_TSnw}\right)+\frac{(M-\bar{x}_S+\bar{x}_T)^2}{S_{xx}}\right\}} \qquad (67) \\ &= -0{\cdot}1247+\frac{0{\cdot}0208}{0{\cdot}885} \\ &\qquad -\frac{1{\cdot}96}{11{\cdot}82\times 0{\cdot}885}\sqrt{\left\{0{\cdot}885\times\left(\frac{1}{17{\cdot}5}+\frac{1}{16{\cdot}8}\right)+\frac{(0{\cdot}0208)^2}{0{\cdot}2385}\right\}} \\ &= -0{\cdot}1012-0{\cdot}1874\times\sqrt{0{\cdot}10505} \\ &= -0{\cdot}1012-0{\cdot}0607 \\ &= -0{\cdot}1619. \end{aligned}$$

Similarly, the upper limit is

$$-0{\cdot}1012+0{\cdot}0607 = -0{\cdot}0405.$$

The antilogarithms of these give 0·689, 0·911 as the fiducial

limits to be attached to the statement that the potency of T relative to S is 0·787.

108. An 8-point assay. Simple though the 4-point assay is, it has a number of disadvantages. Its inability to provide adequate validity tests has been noted in §§ 49 and 104, and other shortcomings will be discussed in Section XI. Experimentation on three or more different doses of each preparation is often preferable. Hemmingsen (1933) has reported the figures in Table XXX as the result of an insulin assay by the mouse convulsion method (Chapter V). He used four doses of each preparation, each dose being 1·5 times the next lower dose; he guessed that the test preparation contained about 15 units of insulin per mg., and chose test doses which would be equivalent to those of the standard if this assumed potency were correct. The numbers of mice for corresponding doses were equal, but this is not an essential feature of the assay.

TABLE XXX. *Numbers of mice convulsed in insulin assay*

(Doses in 0·001 unit; those of test preparation calculated from assumed potency of 15 units per mg.)

	Standard preparation No. of mice		*Test preparation No. of mice*	
Dose	*Injected*	*Convulsed*	*Injected*	*Convulsed*
4	12	1	12	1
6	24	8	24	16
9	24	15	24	22
13·5	10	8	10	10

Inspection of Table XXX shows that, except at the lowest dose, the convulsion rate was always higher for the test preparation than for the corresponding dose of the standard: the assumed potency of 15 units per mg. was too low, but probably not very seriously wrong. By application of the methods used above for the strophanthus assay, a better estimate of potency can be formed. The calculations will not be described in detail, as they present few new features.

109. Regression calculations. Table XXXI is of the same pattern as Table XXIX, and contains the calculations needed for estimating the regression equations and making tests of assay validity. Since the doses for each preparation are in

geometrical progression, the log doses will be equally spaced; for arithmetical convenience, they may be taken as 0, 1, 2, 3, providing that the logarithm of the relative potency is multiplied by 0·1761 (equation (31); $0 \cdot 1761 = \log_{10} 1 \cdot 5$) before its antilogarithm is taken. The empirical probits are plotted against x in Fig. 19, in which diagram two parallel lines have

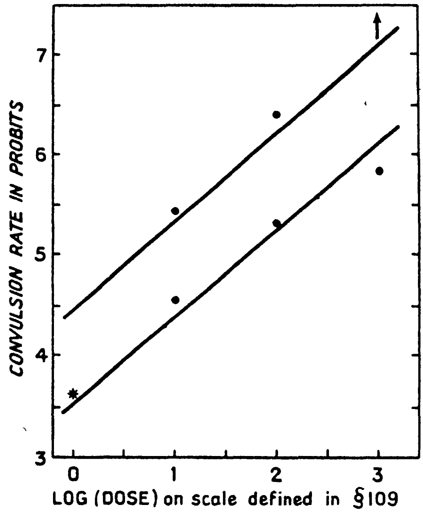


FIG. 19. Assay of insulin by mouse convulsions. (Data in Tables XXX and XXXI.)
Upper line: Standard preparation.
Lower line: Test preparation.
At $x = 0$, both preparations had the same convulsion rate; at $x = 3$, the standard preparation convulsed 100 per cent.

been drawn as provisional regression lines (§ 85), with allowance for a point above the test line at $x = 3$. The horizontal distance between the lines gives

$$M = 1 \cdot 10$$

as a provisional estimate (cf. § 103), which is not in very close agreement with that calculated in § 112, on account of the difficulty of drawing thoroughly satisfactory provisional lines through points as irregular as those in Fig. 19. The expected probits (§ 86) have been read from the provisional lines, to the nearest 0·1. Appendices VII and VIII give the weights and

working probits in the usual manner, interpolation in Appendix VIII being required for certain values (§§ 87, 88).

TABLE XXXI. *Calculations for fitting parallel probit regression lines to the data of Table XXX*

x	n	r	p	*Empirical probit*	*Expected probit*	nw	y	nwx	nwy
Standard preparation									
0	12	1	8	3·59	3·5	3·2	3·61	0·0	11·552
1	24	8	33	4·56	4·4	13·4	4·57	13·4	61·238
2	24	15	62	5·31	5·3	14·8	5·30	29·6	78·440
3	10	8	80	5·84	6·1	4·0	5·80	12·0	23·200
Totals .	.	.	.	.	.	35·4	..	55·0	174·430
Test preparation									
0	12	1	8	3·59	4·4	6·7	3·82	0·0	25·594
1	24	16	67	5·44	5·3	14·8	5·44	14·8	80·512
2	24	22	92	6·41	6·2	8·9	6·38	17·8	56·782
3	10	10	100	..	7·1	1·1	7·51	3·3	8·261
Totals .	.	.	.	.	.	31·5	..	35·9	171·149

$$\bar{x}_S = 1{\cdot}568, \qquad \bar{y}_S = 4{\cdot}927,$$
$$\bar{x}_T = 1{\cdot}140, \qquad \bar{y}_T = 5{\cdot}433.$$

	$Snwx^2$	$Snwxy$	$Snwy$
Standard:	108·600	287·718	871·85
	85·452	271·007	859·49
	23·148	16·711	12·36
			12·06
			$0{\cdot}30 = \chi^2_{[2]}$
Test:	60·300	218·859	960·06
	40·915	195·056	929·90
	19·385	23·803	30·16
			29·23
			$0{\cdot}93 = \chi^2_{[2]}$
Totals:	42·533	40·514	42·52
			38·59
			$3{\cdot}93 = \chi^2_{[5]}$

110. Regression equations. From the totals at the bottom of Table XXXI, the regression coefficient is

$$b = \frac{40{\cdot}514}{42{\cdot}533}$$
$$= 0{\cdot}9525.$$

The regression equations are then found as

$$Y_S = 3{\cdot}433 + 0{\cdot}952x,$$

$$Y_T = 4{\cdot}347 + 0{\cdot}952x,$$

the calculations here being of the same nature as in § 105. Evaluation of each equation for $x = 0, 1, 2$, and 3 shows good agreement with the expected probits used in Table XXXI, and no second cycle of computation is needed.

111. Validity tests. The sums of squares and products of deviations for each preparation have 3 degrees of freedom, and therefore for each the calculation

$$S_{yy} - \frac{(S_{xy})^2}{S_{xx}}$$

gives a $\chi^2_{[2]}$ for a test of the significance of deviations from linearity (§§ 93, 104). Because of the additive property of the statistic χ^2, the sum of these gives a linearity test for the two preparations together:

$$\chi^2_{[4]} = 0{\cdot}30 + 0{\cdot}93$$
$$= 1{\cdot}23,$$

the degrees of freedom for this χ^2 being the sum of the degrees of freedom for its constituents. Reference to Appendix V shows the χ^2 to be non-significant.

The pooled sums of squares and products at the bottom of Table XXXI have 6 degrees of freedom; a χ^2 calculated from them will therefore have 5 degrees of freedom:

$$\chi^2_{[5]} = 42{\cdot}52 - \frac{(40{\cdot}514)^2}{42{\cdot}533}$$
$$= 3{\cdot}93.$$

This χ^2 is a measure of the total deviations of the data from the calculated parallel regression lines, and must include both deviations from linearity for lines fitted separately to the two preparations, and the deviation from parallelism. A test of parallelism is given by subtraction of the linearity component from the total, to leave a χ^2 with 1 degree of freedom:

$$\chi^2_{[1]} = 3{\cdot}93 - 1{\cdot}23$$
$$= 2{\cdot}70.$$

There is seen to be no significant evidence that the true regression lines are not parallel.

112. Estimation of relative potency. By equation (32), as in § 106,

$$M = 1{\cdot}568 - 1{\cdot}140 - \frac{4{\cdot}927 - 5{\cdot}433}{0{\cdot}9525}$$
$$= 0{\cdot}959.$$

An arbitrary scale of log doses has been used, and, as stated in § 109, M must be multiplied by 0·1761 to convert it into $\log R$ (cf. §§ 48, 51). Hence

$$\log R = 0{\cdot}1689$$

and $$R = 1{\cdot}475.$$

Since all the calculations have been based upon an assumed potency of 15 units per mg., the potency of the test preparation is now estimated as 15 × 1·475 or 22·1 units per mg.

113. Precision of potency estimate. Determination of fiducial limits by the approximate formula for $V(M)$ should be contingent upon g being less than 0·1. From equation (64)

$$g = \frac{(1{\cdot}96)^2}{(0{\cdot}952)^2 \times 42{\cdot}53}$$
$$= 0{\cdot}100,$$

so that use of $V(M)$ is barely justified. The assay provides an interesting opportunity for comparing the true fiducial limits with those derived from $V(M)$. The approximate variance formula, equation (66), gives

$$V(M) = \frac{1}{(0{\cdot}9525)^2} \times \left\{\frac{1}{35{\cdot}4} + \frac{1}{31{\cdot}5} + \frac{(0{\cdot}531)^2}{42{\cdot}53}\right\}$$
$$= 0{\cdot}0734.$$

Hence the standard error of M is assessed as 0·271; multiplication by 1·96 gives 0·531 as the distance to the fiducial limits, which are therefore placed at 0·428 and 1·490. These lead to 17·8 units per mg. and 27·4 units per mg. as the limits for the potency.

On the other hand, the formula (67) gives for the lower limit to M

$$0{\cdot}428+\frac{0{\cdot}531}{0{\cdot}900}$$
$$-\frac{1{\cdot}96}{0{\cdot}9525\times 0{\cdot}900}\sqrt{\left\{0{\cdot}900\times\left(\frac{1}{35{\cdot}4}+\frac{1}{31{\cdot}5}\right)+\frac{(0{\cdot}531)^2}{42{\cdot}53}\right\}}$$
$$= 1{\cdot}018-0{\cdot}563$$
$$= 0{\cdot}455;$$

similarly, the upper limit is (1·018+0·563) or 1·581. These give a lower limit of 18·0 units per mg. and an upper of 28·5 units per mg. for the potency of the test preparation. The discrepancy in the upper limit given by the approximate method is perhaps not very serious, but it is large enough to illustrate the dangers of using the approximation. A smaller value of g would have brought the two sets of limits into closer agreement, but a larger value might make the discrepancy so great that the approximate method was worthless.

114.† Heterogeneity. If the χ^2 obtained for the test of linearity is significant, but inspection of a diagram shows the points to be widely scattered about the regression lines rather than having any curvilinear trend, the cause may be heterogeneity of the batches of subjects, and not invalidity of the analysis. As described in § 94, a mean square obtained by division of χ^2 by its degrees of freedom should then be taken as a variance, s^2. The test of parallelism should be made in the form of a variance ratio test, the mean square for deviations from parallelism being compared with s^2. All variances, g, and the whole of the square root term in the fiducial limit formula (67) should be multiplied by s^2, and t should be assigned the number of degrees of freedom on which s^2 is based. Most users of the probit method will be well advised to seek help from a statistician before adopting this procedure, which may give misleading results if applied uncritically.

XI. The Design of Assays

115. General principles. The prime object of any biological assay is:

(i) To obtain a valid estimate of the potency of a test preparation relative to that of a standard.

This in itself is of little use unless the investigator is also able:

(ii) To ensure that the potency estimate is of reasonably high precision.

A vague assurance that the estimate is reasonably precise is seldom sufficient, and usually the assay must be planned:

(iii) To provide a numerical assessment of the precision of the estimate.

Indeed (ii) is rarely accomplished without (iii), though the two are logically distinct. Unless belief in the validity of the method *a priori* is so strong as to make the precaution unnecessary, the investigator will wish:

(iv) To include a test, or tests, of the validity of the assay and of the statistical analysis applied to the data in order to obtain the potency estimate (§ 45).

By a valid assay is meant one which genuinely estimates the quantity it is intended to measure, and fails to give the true value only because of the unavoidable variations between test subjects. A valid estimate will not be obtained unless both the biological technique is capable of giving results relevant to the potency whose estimate is wanted, and the statistical technique is appropriate to the dose-response relationship encountered.

If the test preparation is known to be chemically identical with the standard, except for a different amount of inert diluent, then any biological technique for establishing equally effective doses of the two must lead to a valid estimate of relative potency. The species of the subject, the nature of the response, and the conditions of the experiment, will affect the precision of estimation, but all the alternatives will give valid estimates. On the other hand, if impurities or other constituents of the test preparation may also affect the response, the assay should be designed to give some indication of the occurrence of any disturbance from this cause. Statistical analysis cannot discover the presence of impurities which produce responses whose relationship to dose is of the same kind as that for the substance to be estimated, but, in the parallel line assay, for example, the test of parallelism gives some hope of detecting disturbances in the character of the dose-response relationship. If the active principles of the two preparations are not chemically identical,

the necessity of checking the parallelism of the curves relating response to log dose before using a relative potency figure is still greater. The methods of statistical analysis of biological assay data described in this chapter depend upon representation of the dose-response relationship (possibly in terms of metameters other than those directly measured) by a straight line. Though the assay be valid, the analysis will be invalid if the data controvert this linearity, and some other form of analysis must then be sought. For many of the commoner types of assay, experience has taught the right form, but even for these, and still more for newer methods of standardization, linearity tests are a valuable safeguard against the misuse of statistics.

The *design* of an assay, that is to say, the choice of the levels of dose to be used and of the subjects to be assigned to each dose, involves compromise between precision of estimation and sensitivity of validity tests. To satisfy requirements (ii) and (iii) simultaneously is not difficult, provided that the number of subjects available is not too small, but in order also to obtain reasonably sensitive validity tests some sacrifice of precision is unavoidable. Failure to realize this conflict of interests is responsible for many bad assay designs. Fortunately, as will be seen from later paragraphs, the sacrifice need not be great, particularly if a potency estimate not too wildly in error can be guessed at the start of the assay.

116. Pilot investigations. Some knowledge of the form of a dose-response relationship should always precede any attempt to use that relationship as the basis of an assay. Before any newly investigated dose-response relationship is used for assay purposes, special experimentation on a standard preparation is needed in order that the form of the relationship may be discovered. Particularly important is knowledge of whether responses have a linear regression on dose or on log dose, or whether some more complex linearizing transformation is needed, and over what range of doses any regression remains linear.

When this information has been obtained, the investigator is in a position to begin assaying test preparations relative to the standard. As will be made clear in subsequent paragraphs, the planning of a good assay is greatly assisted by approximate knowledge of the potency to be estimated. Occasionally, before a full-scale assay of a new preparation (whose potency is quite

unknown) is undertaken, a small or *pilot assay*, using very few subjects, may be performed in order to indicate the order of magnitude of R. Not only will the pilot assay help the investigator to choose suitable doses of the test preparation for use in the main assay, but it will also give approximate values of s and b which, when inserted in the appropriate formula for $V(M)$, will suggest the number of subjects needed in the main assay if a specified precision is to be obtained. Of course, a pilot assay is not always desirable; in situations where it is thought likely to be valuable, expert statistical advice on the most economical procedure and the use to be made of the results will probably be wanted.

117. Symmetry of design. In general, fully symmetrical assay designs are to be recommended, both for efficiency and for convenience. That is to say, there should be equal numbers of doses for the two preparations, evenly spaced on the dose scale (usually absolute or logarithmic) to be used for linearity, and equal division of the total number of subjects between all the doses. Equal division of effort between the standard and the test preparation is desirable for maximizing the precision, and the other symmetries are practical conveniences both in experiment and in computation, which should be maintained unless there are good reasons to the contrary. Of course, perfect balance may be upset by the necessity of rejecting some subjects or some doses (but see Chapter II on 'Discordant results').

118. Choice of doses. The doses used in an assay should always extend over the widest range possible without going beyond the region of linearity, as this leads to the most reliable estimates of the regression coefficients. The range, however, should err by being too narrow rather than too wide, and, since the potency of the test preparation is not known in advance, some contraction may be required because of the convenience of using equal ranges for the two preparations. Points which are shown by the dose-response diagram as 'obviously' outside the linear region may sometimes be discarded from the analysis and potency estimation, but care is needed if the word 'obviously' is not to become an excuse, possibly subconscious, for tampering with the results in such a way as to make them agree with the investigator's preconceived ideas.

119. Number and grouping of subjects. Potency estimates of

satisfactory precision are unlikely to be obtained unless at least twenty subjects are employed, and more will be needed if the response measured varies widely from one individual to another. In assays based upon quantal responses, two to three times as many subjects should be used, because of the less complete information given by each. The actual number chosen is a matter for special consideration in any given type of assay, in the light of any evidence on variability that may be available (§ 116).

Whenever differences between litters, or between other groupings of the test subjects, are likely to be large, an attempt should be made to incorporate some form of litter-mate control (§ 30), or other block restriction, into the assay. The group differences are thereby balanced over the dose-levels, so that they do not affect the precision of the potency estimate. Even though the number of doses to be tested exceeds the number of subjects per litter, it may be possible to devise some measure of control of this source of variation; on these *incomplete block* designs, the advice of a statistician will be needed (Bliss, 1947).

120. Detailed recommendations on design. In the following paragraphs, detailed consideration is given to the problem of the planning of an assay with a limited number of subjects. It is assumed that the pilot investigations have been completed, so that the investigator has some idea of the values of R, s, and b—though he must remember that any change in experimental conditions or in the source of his test subjects may affect s and b. By use of a sufficiently large number of subjects, even an assay of poor design may be made to yield a potency estimate of any desired precision. In practice, however, the investigator is concerned to make the most effective use of a maximum of N subjects, or even to decide whether, without exceeding this number, any assay of worth-while precision is possible.

The more mathematical aspects of this problem have been described elsewhere (Finney, 1947*b*), and can only be summarized here. Special recommendations for parallel line assays (Sections V and VI) are made in §§ 121, 122. These are followed by several paragraphs on cross-over designs. Corresponding recommendations for slope-ratio assays (Section VII) are given in §§ 128, 129, and for quantal response assays (Section X) in § 130.

121.† Parallel line assays. If the investigator is able to use N

subjects for the assay, he will assign $\frac{1}{2}N$ to each preparation in order to have a symmetrical design (§ 117). The next question to be decided is how many dose-levels of each are to be tested. The general symmetrical design may be described as $2k$-point, having k doses of each preparation at equal intervals of x and with a range d (in logarithmic units) between the highest and lowest doses of either preparation; d is chosen as large as possible, consistent with the requirement of keeping within the linear portions of the response curves (§ 118).

Provided that g is sufficiently small, fiducial limits to M, the log relative potency, will be derivable from the formula for $V(M)$ given as equation (33). For a regular spacing of doses, a little algebra shows that

$$S_{xx} = \frac{Nd^2(k+1)}{12(k-1)};$$

substitution of this and $N_S = N_T = \frac{1}{2}N$ in equation (33) gives

$$V(M) = \frac{4s^2}{Nb^2}\left\{1+\frac{3(k-1)(M-\bar{x}_S+\bar{x}_T)^2}{d^2(k+1)}\right\} \qquad (68)$$

(Finney, 1947*b*, equation (67)). Now s and b are characteristics of the experimental technique and the stock of test subjects, and cannot be affected by the set of doses chosen for a particular assay. In terms of quantities which can be controlled by suitable choice of design, therefore, the interval between the fiducial limits of M is seen to be proportional to

$$\sqrt{\left(1+\frac{3(k-1)}{k+1}u^2\right)}, \qquad (69)$$

where u is defined by

$$u = (M-\bar{x}_S+\bar{x}_T)/d, \qquad (70)$$

the ratio of the discrepancy between M and the mean dose difference to the total dose range. Consequently, if $u = 0$ the interval between the fiducial limits is independent of k, but otherwise it increases with k. The behaviour for various values of u and k is shown in Fig. 20, in which the percentage increase in the fiducial range (as compared with $u = 0$) is plotted against k. In practice, unless the test preparation is something entirely new, it should be possible so to choose doses that u does not exceed 0·3 in absolute value (that is to say, u is neither greater than 0·3

nor less than $-0{\cdot}3$).[1] Use of 5 or 6 doses, instead of 2, should not then increase the fiducial range of M by more than 10 per cent.

The condition that g should be small is important. An increase in the number of doses from 2 to k will have the effect of multiplying g by $3(k-1)/(k+1)$, because of the lesser precision in determining the regression coefficient, and so may convert a value of g less than $0{\cdot}1$ into one that can no longer be

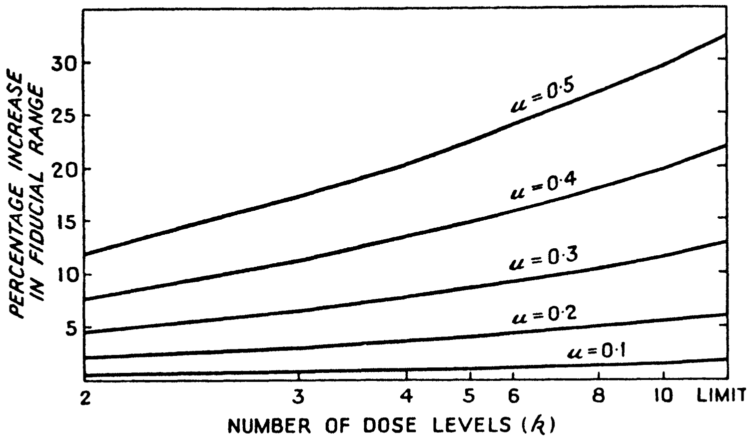


FIG. 20. Percentage increase in the fiducial range for the estimate of log potency in a parallel line assay, for different numbers of dose levels, as compared with 4-point assay having $u = 0$. (See § 121.)

ignored: a 10-point assay ($k = 5$), for example, will have g twice as large as a 4-point assay with the same interval between the highest and lowest doses. When g is not negligible, the loss of precision in the estimation of M on account of increase in k is more serious than Fig. 20 indicates (Finney, 1947*b*, Appendix).

122. Recommendations for parallel line assays. The variance of M will be minimized, and the fiducial limits brought as close as possible to M, if the doses can be so chosen that

$$\bar{x}_S - \bar{x}_T = M.$$

[1] From § 51, for the oestrone assay

$$M - \bar{x}_S + \bar{x}_T = -0{\cdot}188,$$

and, in the dose scale explained in § 48, $d = 2$. Hence the absolute value of u is less than $0{\cdot}1$. Similarly, in the assay of testosterone propionate, from § 58

$$M - \bar{x}_S + \bar{x}_T = 0{\cdot}158,$$

and again $d = 2$. Thus both these assays satisfy the condition, though the first has too large a value of g for the present argument to apply.

This, of course, cannot be accomplished deliberately, for if M were known the assay would not be needed! Nevertheless, from a pilot assay or otherwise, the investigator usually has some idea of a likely value for the relative potency and hence for M, its logarithm, and he should take account of this information in the selection of his doses. He should first choose log doses for the standard preparation spaced at equal intervals over what he believes to be the extent of the linear response law. If he then forms a set of doses for the test preparation by subtraction of a guessed value of M from each x_S, he will have

$$\bar{x}_T = \bar{x}_S - M,$$

and, if his guess at M is good, he will nearly satisfy the condition for minimizing $V(M)$.

The most precise estimate of M would be given by a 4-point assay, using $N/4$ subjects (where N is the total number available) at the highest and lowest doses of each preparation compatible with linearity. This, however, fails to give satisfactory validity tests (§§ 45, 49). There is no independent test of linearity, and the parallelism test is unlikely to be sensitive to deviations from linearity. In spite of its simplicity and convenience, the design cannot be recommended except for routine assays of materials whose behaviour is so well known that no doubts of the validity of either the biological procedure or the statistical analysis exist: even complete non-parallelism of response curves might escape detection in a 4-point assay.

The loss in precision that results from using a 6-point design instead of a 4-point should be small (Fig. 20): probably the best general purpose design is that employing $N/6$ subjects at each of three doses of the two preparations, taking an additional dose half-way between the highest and lowest (on the logarithmic scale). Validity tests can then be made as in § 57. An 8-point or a 10-point design can also be used without much loss of precision, and will give more information on validity, but these may be unduly complicated for most purposes.

The precision of estimation of relative potency will also be increased by any procedure which reduces s or which increases the regression coefficient, b. Selective breeding, aimed at increasing the sensitivity of subjects to small changes of dose as well as at reducing the variability in their responses to a fixed

dose, and careful choice of experimental conditions, may be important in this connexion.

123. Cross-over designs. In all the earlier discussions of assays, it has been assumed that only one response could be determined from one subject. This is necessarily so if measurement of the response entails dissection of the animal. Even when the animal does not have to be killed before the response can be measured, use of the same subject on two or more occasions might be unsafe because the response to the second dose was not independent of the first dose received. Nevertheless, since in all biological assays the variability between subjects is one of the chief reasons for low precision in potency estimates, a procedure which allowed responses to both preparations to be measured on the same subject would be welcomed whenever the fear of non-independence of successive responses was absent.

Various arrangements known as *cross-over designs* have been suggested, in which each subject is made to give two response measurements, and differences between subjects can be eliminated as a cause of error in M. These designs have been found particularly valuable for insulin assays using the percentage fall in the blood-sugar of injected rabbits (Chapter V). The *twin cross-over design* is developed from the 4-point. Two doses of each preparation, denoted by S_1, S_2, T_1, T_2, are required, and are assigned to a group of four rabbits in the following order:

	Dose on	
Rabbit no.	*Day I*	*Day II*
1	S_1	T_2
2	S_2	T_1
3	T_1	S_2
4	T_2	S_1

The treatments are repeated on further groups of four rabbits. Comparisons between tests on the same rabbit can then be used to estimate the difference in response for doses S_1 and T_2, and for doses T_1 and S_2; the sum of these gives the average difference between high and low doses, from which the regression coefficient is obtained, and the difference gives the average difference in response between the two preparations. Balance over the two days of the experiment, on the assumption that

responses on Day II are independent of what happened on Day I, eliminates any effect of differences in responsiveness on the two days.

The design of an experiment determines the pattern of the statistical analysis, and the twin cross-over cannot be analysed as though it were an ordinary 4-point. The estimation of relative potency follows the usual scheme, and presents no difficulty unless missing observations cause a complication. The analysis of variance, the calculation of s^2, and the validity test, however, require modification of the procedure described in §§ 48–52; an example is described in §§ 124, 125.

If the test subjects can be used more than twice, further permutations of the doses may be introduced. Of particular interest are designs in which each animal receives every dose of both preparations; for example, if the rabbits of the insulin assay could be used four times each, the first set of four rabbits might have been treated as follows:

	Dose on			
Rabbit no.	*Day I*	*Day II*	*Day III*	*Day IV*
1	S_1	T_2	T_1	S_2
2	S_2	T_1	S_1	T_2
3	T_1	S_2	T_2	S_1
4	T_2	S_1	S_2	T_1

Bliss and Marks (1939) have described these *Latin square* designs. The objections previously raised to 4-point assays apply with equal force to cross-over designs involving only two doses of each preparation. The cross-over principle may be extended to assays based on more doses: a 6-point assay might be arranged as:

	Dose on	
Rabbit no.	*Day I*	*Day II*
1	S_1	T_3
2	S_2	T_2
3	S_3	T_1
4	T_1	S_3
5	T_2	S_2
6	T_3	S_1

with repetitions on other groups of six rabbits. The Latin square scheme might be adopted if each rabbit could be used six times. It is impossible to give here a full account of all available designs of this type and of their statistical analyses.

124. A twin cross-over assay. Smith, Marks, Fieller, and Broom (1944) give the results of an assay of insulin by the twin cross-over method. Doses of the standard preparation were 1 and 2 units per ml., and those of the test preparation were equivalents calculated from an assumed potency of 22 units per mg. for the sample. Twelve rabbits were used, and the second day of testing was one week after the first. The responses, shown in Table XXXII, were percentage reductions in blood-sugar adjusted for the initial level (see Smith *et al.* for details).[1]

TABLE XXXII. *Responses in a twin cross-over insulin assay*

Rabbit no.	*Day I*		*Day II*		*I+II*
1	47	S_1	52	T_2	99
2	48		46		94
3	38		34		72
Totals . . .	133		132		265
4	38	S_2	26	T_1	64
5	53		34		87
6	38		29		67
Totals . . .	129		89		218
7	34	T_1	34	S_2	68
8	26		24		50
9	43		43		86
Totals . . .	103		101		204
10	34	T_2	22	S_1	56
11	40		25		65
12	39		18		57
Totals . . .	113		65		178
	478		387		865

[1] In this assay, records for one rabbit, no. 7, are missing; for the purpose of illustration here, values have been inserted and used as if they were genuine: the exact analysis introduces complications beyond the scope of this chapter.

The mean responses are

$$\bar{y}_S = \frac{133+129+101+65}{12}$$
$$= 35{\cdot}67$$

and
$$\bar{y}_T = \frac{132+89+103+113}{12}$$
$$= 36{\cdot}42.$$

If the log doses are taken as -1, $+1$ for each preparation, as in § 48, the average regression coefficient is

$$b = \frac{-133+132+129-89-103+101+113-65}{24}$$
$$= \frac{85}{24}$$
$$= 3{\cdot}542.$$

Hence, provided that the assay is valid—and this question will be discussed below—an estimate of relative potency can be derived from equation (32) as

$$M = \frac{36{\cdot}42-35{\cdot}67}{3{\cdot}542}$$
$$= 0{\cdot}212.$$

By equation (31), this must be multiplied by 0·1505 ($= \frac{1}{2}\log 2$) to give the logarithm of R, and therefore

$$R = 1{\cdot}076.$$

The assumed potency was 22 units per mg. and multiplication by R gives 23·7 units per mg. as the estimate.

This estimation follows exactly the same course as that in § 51. The figure obtained is almost certainly more precise than would have been given by an ordinary 4-point assay with six rabbits at each dose used once only; without a full analysis of variance, however, it is impossible to judge whether the assay is valid and what its precision is. This analysis presents no arithmetical difficulties, but those unpractised in the analysis of variance will find it complicated. Considerable experience is needed before the various independent components of the analysis can be identified and put into the right places with

confidence. The account that follows is necessarily brief, and further guidance on the technique should be sought in the books and papers already quoted.

125.† Analysis of variance. Table XXXIII is the analysis of variance of the data in Table XXXII. The adjustment for the mean,

$$\frac{865^2}{24} = 31{,}176{\cdot}04,$$

is first calculated (cf. § 25 and other analyses of variance), and entered at the top of the table. The total sum of squares is next formed, in the usual manner. From it is removed a portion, with 11 degrees of freedom, representing differences between rabbits; if R_1, R_2, R_3,..., R_{12} represent the totals of the two responses for each rabbit, this sum of squares is calculated from formula (18) as

$$\frac{R_1^2+R_2^2+R_3^2+...+R_{12}^2}{2}-\frac{G^2}{24}$$

$$= \tfrac{1}{2}\times(99^2+94^2+...+57^2)-31{,}176{\cdot}04$$

$$= 1{,}376{\cdot}46.$$

The remainder, after subtraction of this from the total, 722·50, is a sum of squares for variation 'within rabbits', with 12 degrees of freedom, which could be separately evaluated by summation of the twelve components $\tfrac{1}{2}(52-47)^2$, $\tfrac{1}{2}(46-48)^2$, &c.

The difference between preparations is formed from 'within rabbit' comparisons, and in the usual way (§§ 48, 56) gives a sum of squares

$$\frac{(132-133+89-129-101+103-65+113)^2}{24} = 3{\cdot}38.$$

Similarly the regression sum of squares is formed from differences between levels, and again one of the two responses for each rabbit occurs with a positive sign, the other with a negative: the overall difference has already been found, and the result is

$$\frac{85^2}{24} = 301{\cdot}04.$$

A third component of the 'within rabbits' sum of squares is that

representing the average differences between the two days of experimentation:

$$\frac{(478-387)^2}{24} = 345\cdot04.$$

The remainder, with 9 degrees of freedom, leads to a mean square which will eventually be used as s^2.

TABLE XXXIII. *Analysis of variance for the data of Table XXXII*

Adjustment for mean . . .		31,176·04	
Nature of variation	*d.f.*	*Sum of squares*	*Mean square*
Parallelism . . .	1	18·38	18·38
Days × preparations . .	1	425·04	
Days × regression . .	1	222·04	
Error between rabbits .	8	711·00	88·88
Total between rabbits .	11	1,376·46	
Preparations . . .	1	3·38	
Regression . . .	1	301·04	
Days	1	345·04	
Error within rabbits . .	9	73·04	8·116
Total	23	2,098·96	

The 'parallelism' component may be formed exactly as in § 48 from differences of the type $(T_2-T_1-S_2+S_1)$. Here S_1, T_2 both have positive signs, S_2, T_1 both negative; inspection of Table XXXII shows that each rabbit has one of these pairs of doses, so that the component involves differences between rabbits. The sum of squares

$$\frac{(265-218-204+178)^2}{24} = 18\cdot38$$

must therefore be put in the upper part of Table XXXIII. A second component requiring to be removed from the sum of squares between rabbits is one representing the *interaction* of days and preparations, the extent to which the difference between preparations is greater on one day than on the other. The differences on the two days are

$$\text{I: } -133-129+103+113,$$

$$\text{II: } 132+89-101-65,$$

and subtraction of the first from the second leaves a contrast (§ 56) dependent only on rabbit totals, namely

$$265+218-204-178.$$

The sum of squares, by the usual rules, is

$$\frac{(265+218-204-178)^2}{24} = 425{\cdot}04.$$

A third component is the analogous interaction of days and levels, the extent to which the regression coefficient is greater on one day than on the other; this may easily be shown to give the sum of squares

$$\frac{(265-218+204-178)^2}{24} = 222{\cdot}04.$$

The last two are of no great interest in themselves, but must be deducted from the sum of squares between rabbits before an error mean square for this part of the table is derived.

Had the 12 rabbits been divided into 3 litters of 4, the 4 sequences of doses would have been assigned to 1 rabbit from each litter. A sum of squares, representing differences between litters, with 2 degrees of freedom, would then have been inserted in the 'between rabbits' part of the analysis of variance, and the error sum of squares would have been reduced to 6 degrees of freedom.

126. Validity test. As for the ordinary 4-point assay (§ 49), no test of linearity is available, and the only validity test that can be applied is that for parallelism. The variance ratio test shows that the parallelism mean square is not significantly greater than the 'between rabbits' error (88·88). One of the disadvantages of the design is that this test is not very sensitive, by comparison with the precision of the potency estimation, since it depends upon comparisons between rabbits, whereas M is based upon comparisons within rabbits. In this instance, however, there is no cause for alarm, since the parallelism mean square is, in fact, not much greater than the 'within rabbits' error (8·116).

127. Precision of assay. Assessment of the precision of M proceeds as in § 52. Both numerator and denominator of M

are based on contrasts within rabbits, and therefore s^2 is taken from the lower half of Table XXXIII:

$$s^2 = 8{\cdot}116.$$

From equation (34), with substitution of a t based on 9 degrees of freedom,

$$g = \frac{(2{\cdot}26)^2 \times 8{\cdot}116}{(3{\cdot}542)^2 \times 24}$$
$$= 0{\cdot}138,$$

too large for the use of $V(M)$. Since $\bar{x}_S = \bar{x}_T = 0$, formula (35) for the lower limit to M simplifies to

$$\frac{0{\cdot}212}{0{\cdot}862} - \frac{2{\cdot}26 \times 2{\cdot}849}{3{\cdot}542 \times 0{\cdot}862}\sqrt{\left\{0{\cdot}862 \times \left(\frac{1}{12}+\frac{1}{12}\right)+\frac{(0{\cdot}212)^2}{24}\right\}}$$
$$= 0{\cdot}246 - 0{\cdot}804$$
$$= -0{\cdot}558,$$

and the upper limit is similarly $(0{\cdot}246+0{\cdot}804)$, or $1{\cdot}050$. Multiplication by $0{\cdot}1505$ gives the limits to $\log R$; the limits to R are therefore 0·824, 1·439. Hence the test preparation is estimated to contain 23·7 units insulin per mg., with fiducial limits at 18·1 units per mg. and 31·7 units per mg.

The mean square for error between rabbits in Table XXXIII is almost 11 times that within rabbits, and this is an indication of the great improvement in precision resulting from use of the cross-over design. These figures may be used to estimate that in an ordinary 4-point assay based upon 24 responses, involving one response only from each of 24 rabbits, the value of s^2 would have been about 48·5. Without any allowance for the increase in g, this sixfold increase in s^2 would extend the fiducial range to almost 2·5 times its present width, and the effect on g would probably be such as to make the assay hopelessly imprecise. An alternative form of this statement is that, in a 4-point assay with only one response measure for each subject, about 144 rabbits would be needed to give the same precision.

128.† Slope-ratio assays. The general symmetrical design may be described as $(2k+1)$-point, having k doses of each preparation, together with a blank. The investigator will first choose X_S, X_T doses of the two preparations guessed to be about equipotent, and as large as possible without fear of being beyond the range of linearity. The equally spaced doses will

then be zero, X_S/k, $2X_S/k, \ldots$, $(k-1)X_S/k$, X_S of the standard, and X_T/k, $2X_T/k, \ldots$, $(k-1)X_T/k$, X_T of the test preparation; $N/(2k+1)$ subjects will be assigned to each dose-level.

Provided that g is sufficiently small, equation (43) may be used for $V(R)$, and fiducial limits derived from it. Algebraic analysis similar to that in § 121 can be used to give an expression

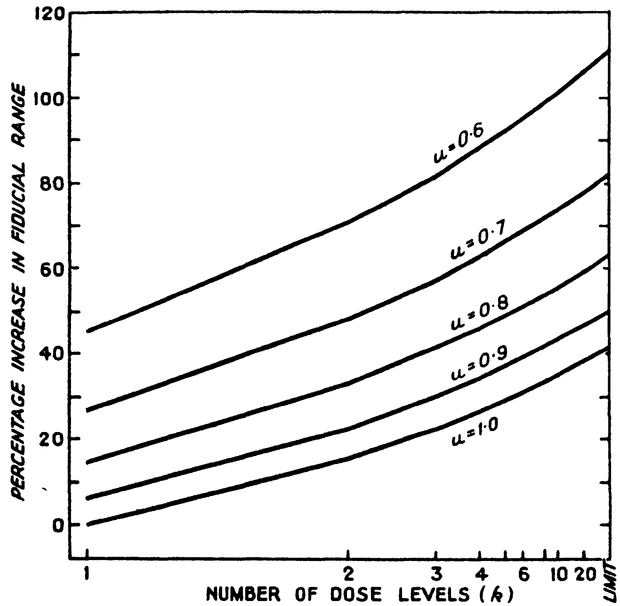


FIG. 21. Percentage increase in the fiducial range for the estimate of potency in a slope-ratio assay, for different numbers of dose levels, as compared with 3-point assay having $u = 1$. (See § 128.)

for $V(R)$ in the symmetrical $(2k+1)$-point assay, a special case of equation (43), as a function of k (Wood and Finney, 1946). From this, it may be seen that the interval between the fiducial limits of R is proportional to

$$\sqrt{\left[\frac{k\{(u^2+1)(5k^2+5k+2)-6uk(k+1)\}}{2(k+1)(k^2+k+1)u^2}\right]}, \tag{71}$$

where $$u = RX_T/X_S, \tag{72}$$

and u is assumed to be not greater than unity (as will generally be true if X_S and X_T have been well chosen). For a 3-point assay, with doses which give $u = 1$, this quantity is unity: its behaviour for other values of u and k is shown in Fig. 21. The

3-point design gives the highest precision, and the loss through increase in k is rather more marked than for parallel line assays. It is perhaps desirable to choose X_T so that u is a little less than unity, in order to avoid the danger of going outside the region of linearity for the test preparation.[1] Use of two doses of each preparation instead of one will increase the fiducial range of R by 15 per cent. of the basic figure for $k = 1$ if $u = 1$, and by 30 per cent. if u falls as low as 0·8. Clearly u should be kept as close to unity as the experimenter dare risk. Further increase of k to 3 or 4 will involve an additional loss of precision, and should not be adopted unless special aspects of validity testing demand it. Wood and Finney have pointed out that use of $N/(2k+1)$ subjects at each dose is not the most efficient scheme for the estimation of R, and that higher precision results from the allocation of more subjects to the larger doses; nevertheless, the practical convenience of the symmetrical arrangement, and its advantages for tests of validity, usually outweigh this consideration.

The condition that g should be small is important. An increase in the number of doses per subject from 1 to k has the effect of multiplying g by

$$\frac{k(5k^2+5k+2)}{2(k+1)(k^2+k+1)},$$

because the placing of subjects at intermediate instead of extreme doses reduces the precision of b_S. A 5-point design therefore has g 1·5 times as large as a 3-point, and an 11-point has g twice as large. As for parallel line assays, when g is not negligible, the loss of precision in the estimation of R on account of increase in k is more serious than Fig. 21 indicates (Finney, 1947*b*, Appendix).

[1] For the riboflavin assay discussed in Section VII, the maximum doses used were 0·2γ riboflavin and 0·05 gm. malt; the potency of the malt was estimated (§ 65) as 2·73γ per gm. Hence

$$u = 2{\cdot}73 \times 0{\cdot}05 \div 0{\cdot}2$$
$$= 0{\cdot}68.$$

Alternatively, since u is defined in such a way as to make it independent of changes in scales of measurement, in terms of the arbitrary scales introduced in § 61

$$X_S = X_T = 2$$

and

$$R = 0{\cdot}683.$$

Again

$$u = 0{\cdot}68.$$

129. Recommendations for slope-ratio assays. If the investigator has some idea of the relative potency, R, he should choose X_S, *the highest dose of the standard preparation, to be as large* as possible without going beyond the linear portion of the response curve, and then take X_T, the highest dose of the test preparation, to be about $0{\cdot}9X_S/R$. He should then divide the N available subjects equally between zero dose and k doses of each preparation at regular intervals from 0 to X_S or X_T (i.e. at X/k, $2X/k$, $3X/k$,...).

The most precise estimate of R would be given by a 3-point assay, using only the doses 0, X_S, X_T. This, however, fails to give any validity tests. The loss in precision through using a 5-point design, with $N/5$ subjects at 0, $\frac{1}{2}X_S$, X_S, $\frac{1}{2}X_T$, X_T, should not be serious (Fig. 21), and this is probably the best general purpose design. Validity tests can then be made as in § 64. A 7-point or a 9-point design gives more information on validity, but entails a greater loss of precision, and is probably unduly complicated for most purposes.

If there is known to be a deviation from linearity at very low as well as at very high doses, the blank may be omitted and a $2k$-point design adopted; the doses will then be equally spaced between the lowest dose of either preparation considered sure to be within the range of linearity and X_S, X_T. Though the design with $k = 2$ will be more precise than that with $k = 3$, the latter is the simplest that can provide validity tests. Wood (1946) has suggested an alternative method of dealing with this situation, which may sometimes be useful.

The precision of estimation of relative potency will be increased by any procedure which reduces s. In contrast with parallel line assays (§ 122), an increase in b_S does not necessarily increase the precision of the estimate. The determining factor is the maximum response that can be obtained on the linear portion of the response curve, and a modification in technique which increases b_S, but merely gives the same maximum response at a lower value of X_S, does not improve the precision. Modification in experimental conditions and selection of test subjects should therefore aim at reducing variability and increasing the response to X_S.

130. Recommendations for quantal response assays. Investigation of the relative merits of different designs for assays based

on quantal responses is more difficult, because of the complications introduced by differential weighting of the observations (§ 87). As for parallel line assays, the log doses should be chosen with a view to having

$$\bar{x}_T = \bar{x}_S - M,$$

but doses giving responses below 5 per cent. or above 95 per cent. are undesirable because of low weight. Again the 4-point design should not be used unless there is practical certainty of validity: indeed, the argument against it is perhaps stronger here than for parallel line assays, because of the danger that unfortunate choice of doses will give 0 per cent. or 100 per cent. response. A symmetrical 6-point design, in which any available information was used to assist the choice of doses of each preparation that would give responses of 15–25 per cent., 45–55 per cent., and 75–85 per cent., should be both of high precision and satisfactory in its validity tests. For study of a characteristic curve (§ 101), of course, far more than three doses of a preparation are needed, but that would be an experimental investigation of the standard preparation rather than itself an assay of relative potency.

A high value of b, which is equivalent to a low standard deviation of individual tolerances, is important for the improvement of precision, and any reasonable steps to increase b should be taken.

131.† Multiple assays. If several test preparations are to be assayed against the same standard, there are two possible procedures. Either each unknown is made the occasion of a separate small assay, or all are included in one *multiple assay*: for example, 2 (or 3) doses of the standard and 2 (or 3) doses of each of 3 unknowns might be used in a single series of tests involving 8 (or 12) dose-levels in all. The multiple assay is generally to be preferred because it entails more economical use of subjects and avoids the extravagance of assigning half the subjects and half the labour to determinations of responses for the standard. If the lengthening in the time or the increase in the number of subjects required for one multiple assay, as compared with one simple assay, leads to much greater heterogeneity amongst the subjects, the advantage may rest with separate assays, though another possibility to be considered is

that of performing two or more multiple assays, each with too few subjects for it alone to give precise estimates, and to average their results.

One arrangement for a multiple assay would be to divide the subjects equally between the standard and each of the test preparations. Examination of formulae expressing the precision of relative potency estimates, however, shows that these estimates will be rather more precise if a greater proportion of the available subjects is assigned to the standard than to any one of the test preparations. The ideal ratio is the square root of the number of test preparations, though slight deviation from this, in order to give convenient whole numbers, causes little loss of efficiency. For example, if 2 preparations are to be assayed simultaneously, subjects should be assigned to the standard and test preparations in the ratio 1·4 : 1 : 1, and if 4 are to be assayed simultaneously subjects should be assigned in the ratio 2 : 1 : 1 : 1 : 1. Thus, if 30 subjects were to be used in the assay of 2 preparations, 12 might be assigned to the standard and 9 to each test preparation (a ratio of 1·3 : 1 : 1), and the potency estimates would then be more precise than if 10 subjects were assigned to each of the 3 preparations; the subjects might be divided equally between 3 doses of their respective preparations (§ 122). Unless many preparations are to be assayed simultaneously, the gain from the square root rule is small. Consequently, if block restrictions, for litter-mate control or some similar purpose (§§ 30, 119), were to be introduced, equal division of subjects between preparations would probably be desirable, in order to avoid serious complications.

The analysis of variance can be expanded for use with multiple assays, in order to give a value of s^2 based on all the information provided by the data. The principles of the analysis remain the same, and the most noticeable change is that the number of degrees of freedom for 'preparations' is increased from 1 to the number of test preparations.

132. Randomization. When the number of doses and the number of subjects to be tested at each dose of an assay have been decided, the subjects for each dose-group should be selected at random from all those to be used in the assay. A strictly and objectively random procedure is important if the danger of unconscious purposive selection by the experimenter is to be

avoided. For example, if a 6-point assay is to use 30 animals, 5 should be selected entirely at random for the lowest dose of the standard, 5 more from the remaining 25 for the second dose of the standard, and so on. If the 30 animals consisted of 6 from each of 5 litters, 1 would be selected randomly from each litter for the first dose, 1 more from each litter for the second dose, and so on.

Many investigators like to make their random arrangements with the aid of packs of cards or by drawing numbered disks or slips of paper from a receptacle in which they have been mixed. Writers concerned with the theoretical implications of randomness have frequently drawn attention to the biases that may arise from schemes of this kind, but in practice, providing that reasonable care is taken in the construction of the randomizing device, probably little harm results from them. Nevertheless, it should be realized that, after very little practice in their use, published *tables of random numbers* will be found much more quickly adaptable to most problems: the labour of preparing the slips of paper and mixing them has been done by the compiler of the table, and the experimenter has only to copy the sequence of records appropriate to his need. Fisher and Yates (1948) give such a table, with instructions for its use.

XII. Computation

133. Accuracy. Most experimenters will readily admit that no amount of statistical analysis can repair the deficiencies of ill-planned experimentation and bad laboratory technique. Perhaps few realize how often good experimental work, combined with perfectly sound statistical theory, is marred or even made positively misleading by careless computing. Illegible figures, misplaced decimal points, wrong factors in converting from one scale of measurement to another, and other faults that must be classed as arithmetical blunders, are rightly considered shameful if found in the work of a professional statistician, yet are frequently regarded as no worse than regrettably unavoidable blemishes (or even as pardonable eccentricities!) in the work of a biologist. Arithmetical accuracy, like care in sterilization of apparatus, care of balances, and other routine laboratory practices, is a prerequisite of a good experimenter. Undoubtedly, it can be achieved only by hard work and conscientious attention

to detail, but, in spite of frequently expressed opinions to the contrary, it *can* be so achieved.

The numerical examples in this chapter have been chosen with special regard for their value in introducing the novice to the types of statistical calculation required in biological assay. For the earlier examples, each step of the arithmetic is described in detail, but later the reader is assumed either to be familiar with the principles, or to be able to reproduce the details for himself by analogy with previous examples. The reader to whom statistical science is a new subject is advised to re-work all these examples for himself, and to add to his experience by performing similar calculations on his own data. If he follows the recommended computational routine with care, a little practice should soon not only make him appreciate the purpose of the calculations, but also give him confidence in his arithmetical accuracy.

134. Machinery. A calculating machine of simple pattern, as well as several slide rules and appropriate books of tables, should be standard equipment for any laboratory in which extensive work on biological standardization (or indeed on any form of quantitative biology) is carried out. If it is to be used for short periods by many different people, a hand-operated machine is perhaps best, and least likely to suffer mechanical troubles; electrical operation has many advantages if one person is to use the machine continuously for long. For the general purposes of a laboratory, two machines of simple type are likely to be more valuable than one more complex model, at double the price, intended for the specialized work of a statistical or mathematical department. Anyone familiar with the basic rules of arithmetic can learn to use one of the simpler types of machine in an hour, and will thereafter save at least 50 per cent. of the time and 75 per cent. of the mental strain that he would otherwise expend in his computing. Use of a calculating machine also encourages the systematic arrangement of computations and the routine application of checks. Often a particular result can be reached by several alternative schemes of calculation, the relative merits of which may depend upon the mechanical aids available. The accounts of the numerical examples in this chapter are intended to guide the reader in systematic computing as well as to illustrate statistical

technique. The processes have been described somewhat dogmatically; the author is well aware that alternative schemes of computation may be recommended by other writers and will give exactly the same numerical results, but those shown here have been chosen on a consistent plan as probably the safest and speediest for those who do not spend a great part of their time with a calculating machine.

135. Checks. All calculations should be carefully checked, preferably in a manner which produces the same results by a different sequence of steps. If a particular set of calculations can be checked only by exact repetition, it is desirable that the check consist of independent calculation by a second person. The eventual application of a check must not be used as an excuse for careless original computation: a check should be a demonstration of correctness, not a search for errors.

136. Presentation. A common fault in computing, less serious in its consequences than is inaccuracy, but undesirable because of its wastage of time, is the retention of more digits than the precision of the original measurements justifies. If the total weight of 7 animals, each weighed correctly to the nearest gramme, is 487 gm., the mean should be quoted as 69·6 gm. rather than 69·5714 gm.! Few biological measurements can be made to more than 3 significant[1] digits, and, though up to 6 may reasonably be retained for intermediate calculations, not more than 4 should normally be given in final estimates and tables of mean values. Use of a slide rule is a valuable corrective to the habit of retaining too many digits, and for some types of calculation it is more convenient and more rapid than a calculating machine.

In the preparation of reports and papers, the writer should bear in mind that several simple tables, each containing only a few entries (and these to 2 or 3 significant digits only), are easily read and understood; a single large table, with a multitude of cross-classified entries each of 5 digits, may fail to make any impression on the reader unless he is sufficiently interested to study it in great detail. This is not a counsel that important features of numerical results should be ignored: extensive tabulation, even showing individual measurements, is sometimes desirable. Before presenting any table, however, the

[1] 'Significant' is used here in its arithmetical, not its statistical sense.

writer should ask himself whether, in fact, it either puts important data on permanent record or conveys information to the reader in the most trustworthy and intelligible manner possible. The object of statistical analysis is to extract from a long series of measurements and observations its chief characteristics: this end is defeated if the results are presented in a manner more likely to mislead and confuse the reader than would a direct statement of the original data.

XIII. Symbols

The notation and symbols used by statisticians are not the least variable part of their subject-matter, and no introduction to statistical methods, such as this chapter attempts to provide, can be consistent with all the standard text-books! The writer has tried to follow the most popular usages and at least to maintain internal consistency, but personal prejudices and the shortness of alphabets have prevented complete achievement of these aims. Never to use one letter with two or more different meanings may seem a good rule, but it is surprisingly difficult to follow; where departure from it proved unavoidable, the two meanings are sufficiently distinct to give little risk of confusion.

The following list shows the more important symbols used in the chapter, the paragraphs in which they were first introduced or defined, and a summary of their meaning:

Symbol	*Introduced in paragraph no.*	*Meaning*
d.f.	12	Degrees of freedom.
ED 50	83	Median effective dose, or dose producing response in 50 per cent. of subjects.
LD 50	83	Median lethal dose, or dose killing 50 per cent. of subjects.
S.D.	6	Standard deviation.
S.E.	10	Standard error.
a	62, 82	Constant term of regression equation, usually equal to $(\bar{y}-b\bar{x})$.
b	23, 82	Regression coefficient, equal to S_{xy}/S_{xx} for one independent variate. For multiple regression, see § 62.

Symbol	*Introduced in paragraph no.*	*Meaning*
c	77	Number of columns in a contingency table.
d	17	Difference between two means.
d	121	Difference between highest and lowest log doses in an assay.
F	26	Ratio of two mean squares (see Appendix III).
g	39, 52, 96	Criterion for validity of inferences based on $V(M)$, equal to $t^2V(b)/b^2$. (See also §§ 18, 66).
G	32	Grand total of a set of values of y.
k	121, 128	Number of doses of each preparation in assay.
m	83	Estimated logarithm of ED 50 or LD 50.
M	38, 51, 103	Estimated logarithm of relative potency.
n	8, 69	Number of observations.
n	84	Number of subjects tested at one dose.
n_1, n_2	26	Degrees of freedom for two mean squares in variance ratio test.
N	120	Total number of subjects in an assay.
N_S, N_T	39, 48	Number of subjects for standard and test preparations.
p	73	Proportion of subjects showing characteristic response, r/n.
P	69	Expected or true response rate corresponding to p.
q	73	$1-p$.
Q	69	$1-P$.
r	69	Number of subjects in batch of n showing characteristic response.
r	77	Number of rows in a contingency table.
R	18, 38, 51, 65, 106	Relative potency.
s^2	10	Estimate of variance.
s	10	Estimate of standard deviation, square root of s^2.
s_d, $s_{\bar{x}}$, $s_{\bar{y}}$, etc.	11, 17	Standard error of quantity shown as suffix.
S	8	Indicates summation of a series of quantities of which a typical member is stated.
S_{xx}	23, 89	Sum of squares of deviations of x; $S(x-\bar{x})^2$ or $Snw(x-\bar{x})^2$.

Symbol	Introduced in paragraph no.	Meaning
S_{xy}	23, 89	Sum of products of deviations of x and y; $S(x-\bar{x})(y-\bar{y})$ or $Snw(x-\bar{x})(y-\bar{y})$.
S_{yy}	25, 89	Sum of squares of deviations of y; $S(y-\bar{y})^2$ or $Snw(y-\bar{y})^2$.
t	13	The statistic used in the t-test (see Appendix II).
u	121	$(M-\bar{x}_S+\bar{x}_T)/d$.
u	128	RX_T/X_S.
$V(\ )$	18	Variance of the quantity in ( ).
w	87	Probit weighting coefficient (see Appendix VII).
x	17, 23	Dose or dose metameter.
$\bar{x}$	17, 89	Mean value of x (unweighted or weighted).
X_S, X_T	128	Highest doses of standard and test preparations on linear portion of response curves for slope-ratio assay.
y	8, 23	Response or response metameter.
y	88	By an extension of meaning, used for the working probit (see Appendix VIII).
$\bar{y}$	8, 89	Mean value of y (unweighted or weighted).
$y_1, y_2, y_3, \ldots$	31	Quantities calculated from the data and inserted as 'missing values'.
Y	23, 82	Expected value of y corresponding to specified x in regression equation. Used especially for expected probit.
χ^2	76	The statistic used in the χ^2 test for heterogeneity (see Appendix V). $\chi^2_{[h]}$ denotes 'χ^2 with h degrees of freedom.'
μ	80	The true mean of a distribution, especially of a distribution of tolerances—estimated by m.
σ^2	17, 80	The true variance of a distribution—estimated by s^2.
σ	17, 80	The true standard deviation.
Σ	8	Indicates summation, like S.

Corresponding symbols for different series of data are often distinguished by the addition of subscripts, as $\bar{x}_A$, $\bar{x}_B$ in § 17, or b_S, b_T in § 62. A bar over a symbol, as $\bar{M}$ (read: 'M-bar') in § 40, is often used to indicate a mean value.

In § 32, the letters *c*, *r*, *R* are used with meanings quite distinct from those above, but the context should prevent this and a few other inconsistencies from causing confusion.

REFERENCES

The list that follows is not intended as a complete bibliography of statistical methods for use in biological standardization. It contains only those books and papers to which reference has been made in the chapter. More extensive lists of references have been given by Bliss and Cattell (1943), and Finney (1947*a*, *b*).

Anderson, 1946, *Biometrics*, **2**, 41.
Bliss, 1935, *Ann. appl. Biol.* **22**, 134, 307.
——, 1938, *Quart. J. Pharm. Pharmacol.* **11**, 192.
——, 1940, *J. Amer. Statist. Assoc.* **35**, 498.
——, 1946, *J. Amer. Pharm. Assoc.* **35**, 6.
——, 1947, *Biometrics*, **3**, 69.
Bliss and Cattell, 1943, *Ann. Rev. Physiol.* **5**, 479.
Bliss and Hanson, 1939, *J. Amer. Pharm. Assoc.* **28**, 521.
Bliss and Marks, 1939, *Quart. J. Pharm. Pharmacol.* **12**, 82, 182.
Bliss and Packard, 1941, *Amer. J. Roentgenol.* **46**, 400.
Bliss and Rose, 1940, *Amer. J. Hyg.* **31**, A 79.
British Standards Institution, 1940, *British Standard Method for the Biological Assay of Vitamin D_3 by the Chick Method*, B.S. no. 911.
Bülbring and Burn, 1935, *J. Physiol.* **85**, 320.
De Beer, 1945, *J. Pharmacol.* **85**, 1.
De Lind van Wijngaarden, 1926, *Arch. Exp. Path. Pharmak.* **113**, 40.
Durham, Gaddum, and Marchal, 1929, *Med. Res. Coun. Spec. Rep. Ser.*, no. 128.
Emmens, 1948, *The Principles of Biological Assay.* Chapman and Hall, London.
Fieller, 1940, *J. roy. Statist. Soc. Suppl.* **7**, 1.
——, 1947, *Analyst*, **72**, 37.
Finney, 1945, *Quart. J. Pharm. Pharmacol.* **18**, 77.
——, 1947*a*, *Probit Analysis.* Cambridge University Press, London.
——, 1947*b*, *J. roy. Statist. Soc. Suppl.* **9**, 46.
——, 1948*a*, *J. Hygiene*, **45**, 397.
——, 1948*b*, *Biometrika*, **35**, 145.
Fisher, 1946*a*, *Statistical Methods for Research Workers* (10th edition). Oliver & Boyd, Edinburgh.
——, 1946*b*, *The Design of Experiments* (4th edition). Oliver & Boyd, Edinburgh.
Fisher and Yates, 1948, *Statistical Tables for Agricultural, Biological, and Medical Research* (3rd edition). Oliver & Boyd, Edinburgh.

GADDUM, 1933, *Med. Res. Coun. Spec. Rep. Ser.*, no. 183.
——, 1945, *Nature, Lond.* **156**, 463, 746.
GALLAGHER and KOCH, 1935, *J. Pharmacol.* **55**, 97.
HEMMINGSEN, 1933, *Quart. J. Pharm. Pharmacol.* **6**, 39, 187.
IRWIN, 1937, *J. roy. Statist. Soc. Suppl.* **4**, 1.
——, 1943, *Quart. J. Pharm. Pharmacol.* **16**, 352.
KNUDSEN, BLACKWELL SMITH, VOS, and MCCLOSKY, 1946, *J. Pharmacol.* **86**, 339.
KNUDSEN and RANDALL, 1945, *J. Bact.* **50**, 187.
MAINLAND, 1938, *The Treatment of Clinical and Laboratory Data.* Oliver & Boyd, Edinburgh.
MATHER, 1946, *Statistical Analysis in Biology* (2nd edition). Methuen, London.
PUGSLEY, 1946, *Endocrinology*, **39**, 161.
SCHILD, 1942, *J. Physiol.* **101**, 115.
SMITH, MARKS, FIELLER, and BROOM, 1944, *Quart. J. Pharm. Pharmacol.* **17**, 108.
SNEDECOR, 1946, *Statistical Methods* (4th edition). Iowa State College Press.
TREVAN, 1927, *Proc. Roy. Soc. B*, **101**, 483.
WOOD, 1946, *Analyst*, **71**, 1.
——, 1947, ibid. **72**, 84.
WOOD and FINNEY, 1946, *Quart. J. Pharm. Pharmacol.* **19**, 112.
YATES, 1933, *Emp. J. Exp. Agric.* **1**, 129.
YOUNG and ROMANS, 1948, *Biometrics*, **4**, 122.

CHAPTER IV

PITUITARY (POSTERIOR LOBE) EXTRACT

The Standard. The international standard is a quantity of dry powder prepared from fresh ox posterior lobe material by thorough extraction with acetone to remove water and fat. This preparation was first made by Smith and McClosky (1924), and on the proposal of Voegtlin was adopted as a standard by the Geneva Conference in 1925. The powder has about seven times the activity of an equal weight of fresh material.

According to the recommendations of the conferences held in Geneva in 1925 and in Frankfurt in 1928, any worker was free to prepare a powder by Smith and McClosky's method, and to regard this as international standard powder. The international standard did not consist of one sample of powder deposited in a central institute for distribution. In 1935, however, the Permanent Commission on Biological Standardization decided that it was desirable to have an international standard for reference, and adopted material preserved at the National Institute for Medical Research, London, for this purpose. This material was prepared in 1926 as a British Standard for pituitary (posterior lobe) extract.

Powder similar to the international standard may be made in the following way. Pituitary bodies are collected at the slaughter-house and frozen as soon as possible after removal from the head. For collecting 10 or 20 glands, a thermos-flask may be used containing a mixture of ice and salt around a boiling-tube. The glands are put in the boiling-tube and carried to the laboratory. The separation of the posterior lobes is much easier when the glands are completely frozen. The posterior lobes are dropped into acetone, using 4 c.c. acetone for each lobe, and are left there for 3 hours; they are then cut into small pieces and put in the same volume of fresh acetone; the material is left in this acetone overnight. The acetone is then poured away, and the pieces of gland are dried in an evacuated desiccator over phosphorus pentoxide or calcium chloride. The dried material is then ground in a mortar to pass through a No. 40 sieve. The powder is again dried in a desiccator overnight and extracted

with acetone for 3 hours in a Soxhlet continuous extraction apparatus. It is once more finally dried in a desiccator.

The powder is best kept in a desiccator over phosphorus pentoxide, or in sealed glass tubes which when being prepared are filled with dry air. The tubes to be filled in this way should have a constriction which can be quickly sealed in a blowpipe flame; the appropriate quantity of powder is then put in each, and the tubes are placed upright in a desiccator over phosphorus pentoxide. The desiccator is evacuated and left for 24 hours, and is then slowly filled with air, which passes through a series of three wash-bottles of which the first and third contain concentrated sulphuric acid and the second phosphorus pentoxide, spread on cotton-wool. The tubes are then taken from the desiccator and quickly sealed.

Powder prepared in this way may be of the same strength as the international standard, or it may differ. The international standard is distributed by the National Institute for Medical Research, London; the amount supplied to any laboratory is small, and each laboratory usually prepares a quantity of powder to use as its own sub-standard after comparing it very carefully with the international standard. From five to ten separate comparisons should be made before the relation between the sub-standard and the international standard is regarded as finally determined. The chief cause of variation in the activity of different samples of powder appears to be the variation in the time elapsing between the death of the animal and the freezing of the gland. So far as oxen are concerned, the amount of activity present in glands obtained as freshly as possible varies very little; on the other hand, loss in activity is fairly rapid after the death of the animal, and sometimes it is difficult to discover the interval which has elapsed between death and collection.

The Unit. The international unit of pituitary (posterior lobe) extract is the amount of activity contained in 0·5 mg. of the international standard. This unit is sometimes referred to as the Voegtlin unit, presumably because the use of the powder as an international standard was proposed by Voegtlin. From the definition given it follows that the international standard contains 2,000 units per gramme.

Preparation of an extract. To prepare an extract from the

standard powder, a portion not less than 10 mg. is weighed in a weighing bottle. If the powder is stored in an evacuated desiccator, the desiccator is first slowly filled with air passed through drying bottles, and is then opened. Some of the powder is rapidly transferred to the weighing bottle and the bottle is at once closed. The weighing must not be carried out on a watch-glass as the powder is very hygroscopic; Gaddum (1927) has shown that it gains 10 per cent. of its own weight in 2 minutes. A volume of 0·25 per cent. acetic acid is then taken, so that there is 1 c.c. for each milligramme of powder, and by its aid the powder is transferred to a hard glass boiling-tube. The tube is plugged with cotton-wool and placed in an already boiling water bath for 2 minutes. The tube is then cooled under the tap, the outside wiped dry, and the contents are filtered through a small dry filter-paper into another hard glass tube. The filtrate is a 0·1 per cent. extract of standard, containing 2 units per c.c.

For estimations of oxytocic activity this is diluted ten times with 0·25 per cent. acetic acid and distributed in a series of hard glass tubes which are plugged with cotton-wool and sterilized by placing them in a boiling water-bath for 3 minutes. When cool, these tubes are stored in a refrigerator; the activity will remain unchanged for several months providing the cotton-wool plugs are not removed.

Methods of Estimation

Pituitary (posterior lobe) extract has an oxytocic action, a pressor action, and an antidiuretic action. The antidiuretic method employing rats requires the least apparatus, the least skilled attention, and the least judgement on the part of the operator. To separate the different activities chemically requires complicated treatment which is never used in preparing the usual aqueous extracts, and some workers, therefore, assume that the potency estimated by the antidiuretic method will be the same as that estimated by the oxytocic method. In theory, at least, this may not be so. The oxytocic method has long been established and will be described first.

I. Estimation of Oxytocic Activity

The estimation of oxytocic activity is carried out on uterine muscle. The oldest method is that of Dale and Laidlaw (1912)

which has been used until recently. This employed the uterus of the virgin guinea-pig; it suffered from the disadvantage that suitable animals were not always easy to obtain; and that it often took much time to complete.

Various workers have improved the method, notably Morrell, Allmark, and Bachinski (1940, 1945) and also Hamburger (1945).

The rat uterus has been used instead of the guinea-pig uterus by Bearvillain (1943) and more recently by Mrs. Holton (1948). Her method is as follows.

The rat uterus method

Apparatus. The uterus is suspended in an isolated organ bath of 10 ml. capacity. The bath is filled with Locke's solution which contains half the usual amount of glucose and a quarter of the usual amount of calcium.

Formula for modified Locke's solution

NaCl	45 gm.
KCl	2·1 gm.
$CaCl_2$	0·3 gm.
Glucose	2·5 gm.
$NaHCO_3$	2·5 gm.
Distilled water	5 litres

It was shown by de Jalon, Bayo, and de Jalon (1945) that the rat's uterus has little spontaneous activity in this solution.

A glass tube carries oxygen containing 5 per cent. CO_2 to the bottom of the bath. The pressure of gas should be adjusted so that a stream of fine bubbles is produced. A pin is annealed into the tube near the gas outlet and is used to hold one end of the uterus in position. The other end of the uterus is attached to a length of thread. The thread is fastened to the lever shown in Fig. 22. This lever was designed by Schild (1947) so that the actual shortening of the muscle is magnified in linear proportion on the drum. The lever is also arranged so that the load on the muscle is independent of the inclination of the lever to the horizontal within a certain range. In order to keep within this range stops *EE* are fixed so that the angular deviation is never more than 30° from the horizontal. The lever is made of a strip of aluminium *A*. It is fastened to a wheel *B* which rotates freely about its axle. A long frontal writing point *C* made of glass capillary is attached to *A* by means of a hinge *H*. The diagram

of the lever in Fig. 22 has been drawn to scale and it is particularly important that *A* and *C* are as long as shown in order to produce a faithful record on the drum. A small weight (e.g. a piece of modelling clay) is placed at *F* sufficient to balance the lever when the weight at *P* (which is the point of attachment of the muscle) is 1·3 gm The contractions are magnified about 4

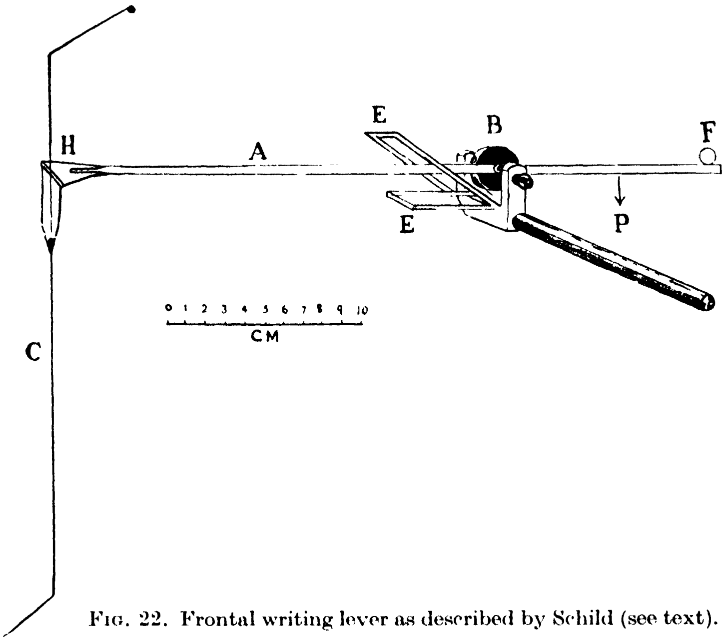


FIG. 22. Frontal writing lever as described by Schild (see text).

times and the maximum excursion of the writing point *D* on the drum is 14 cm. The organ bath is surrounded by a water bath which is kept at a constant temperature (±0·5° C.) of about 32° C. The temperature may be raised or lowered for individual uteri but it is important that once a suitable temperature has been found it should be kept constant throughout the assay.

Choice of the rat. Female rats should be kept separated from the males because pregnant or recently pregnant animals are not suitable. Rats of weights from 140 to 200 gm. have been used with success. In a recent series of tests 8 rats out of 11 were suitable. The uteri of the other 3 had too much spontaneous activity.

Preparation. The rat is killed by a blow on the head and its throat is cut to drain out blood. The uterus is dissected with

scissors and the two horns are divided. One horn is suspended in the bath. The other horn may be kept in Locke solution at room temperature for use later in the day.

Performance of the test. The test is carried out according to Schild's (1942) method for isolated organ assays. It consists of three parts.

I. The range of sensitivity of the uterus to posterior lobe extract must be found by adding doses of the standard preparation. It is best to start by putting 0·02 unit into the bath. After 45 seconds the bath is emptied and refilled with Locke. If 0·02 unit produces a large contraction the next dose should be 0·01 unit. Doses should be given at regular intervals of 2 or 3 minutes and reduced or increased by a multiple of 2 until the range of greatest sensitivity is found. Contractions between 1 and 70 per cent. of the maximal can be used for the assay. Larger contractions should not be used as the relation between height of contraction and log dose is not so nearly linear. It is generally possible to find two doses of standard such that the contractions in response to them are about 10 and 60 per cent. of the maximal. These doses are usually in the ratio 4:3 but may be 3:2 or 8:5. It is obvious that the larger the ratio of 'high' to 'low' dose the less precise will be the result.

Difficulties. (1) The uterus contracts spontaneously. Spontaneous activity may sometimes be overcome by lowering the temperature or by reducing the time interval between successive doses. If spontaneous contractions are troublesome for more than 30 minutes the uterus should be discarded and another rat used. The other horn of the uterus of the first rat will possess the same activity as the first horn. (2) The uterus is very insensitive (i.e. will not contract with 0·1 unit). The sensitivity may sometimes be increased by increasing the temperature or by increasing the time between the doses.

(2). A preliminary estimate of the strength of the unknown solution is made by matching doses of unknown with the two doses of standard. This part of the assay should not take longer than 30 minutes. A dilution of the unknown is then made so that equal volumes of unknown and standard contain what is by this rough estimation the same activity.

(3). The main part of the test consists in recording groups of four contractions, there being in each group a contraction in

response to each of the two selected doses of standard and to each of the two selected doses of unknown. (As already indicated, the ratio of the two doses of unknown must be the same as the ratio of the two doses of standard.) In the different groups the order of the doses must differ. If the greater dose of standard

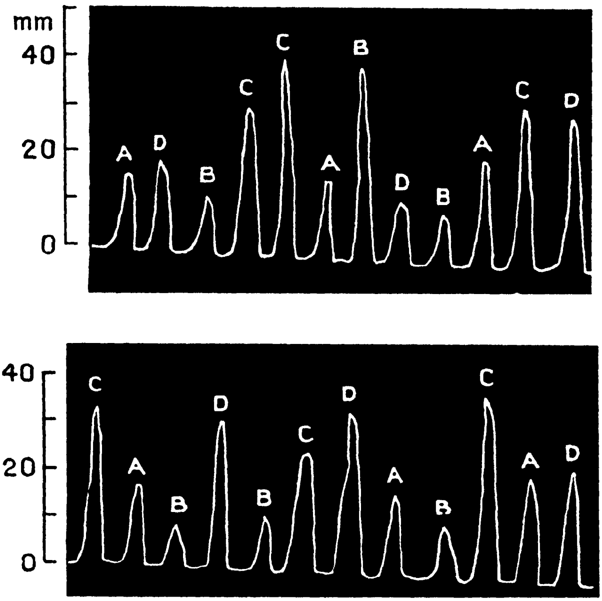


FIG. 23. An assay of pituitary. The record shows 24 contractions of a rat's uterus. The contractions are responses to four different doses of pituitary *A*, *B*, *C*, and *D*, each of which is given once in each group of 4 contractions. There are six groups in all. $A = 0{\cdot}05$ unit, $B = 0{\cdot}04$ unit, $C = 0{\cdot}064$ unit, and $D = 0{\cdot}08$ unit. *B* and *C* were treated as 'standard', *A* and *D* were treated as 'unknown'. $A:D = B:C = 5:8$. Estimate of unknown:standard $= 1{\cdot}25$. True value unknown:standard $= \frac{0{\cdot}05}{0{\cdot}04} = 1{\cdot}25$. A dose was put into the bath every 4 min. and washed out after 45 sec. Weight of rat 140 gm. Temperature 34°–36° C. Load on uterus 1·3 gm. Total experimental time 4 hr. (Record of Dr. Pamela Holton.)

be *A*, the smaller dose *B*, the greater dose of unknown *C*, and the smaller dose of unknown *D*, then Table XXXIV shows the 24 different orders in which the doses can be given. Five or six different groups must be recorded. The response to each dose is then measured in mm., and the result of the test can be calculated. (See Fig. 23.)

Example of a test.

Rat 160 gm. Temperature 31°–33° C.

Time of experiment $3\frac{1}{2}$ hours.

Doses given every 3 minutes. Time for which each dose was allowed to act was 45 seconds.

The two solutions of pituitary extract were:

S = Standard containing 0·5 unit/ml.

U = Unknown believed to contain between 0·3 and 0·8 unit/ml.

1. Preliminary trials with the standard solution showed that satisfactory contractions were obtained with 0·3 ml. and 0·4 ml. of a dilution containing 1·5 ml. of S in 20 ml.

2. A dilution of 1 ml. of U in 10 ml., when given in the same volumes produced contractions about the same as those produced by the diluted standard. Thus the four doses were:

$$0{\cdot}4 \text{ ml. of } \frac{1{\cdot}5\,S}{20} \qquad A$$

$$0{\cdot}3 \text{ ml. of } \frac{1{\cdot}5\,S}{20} \qquad B$$

$$0{\cdot}4 \text{ ml. of } \frac{U}{10} \qquad C$$

$$0{\cdot}3 \text{ ml. of } \frac{U}{10} \qquad D$$

Six groups of four contractions were recorded, the order of doses in consecutive groups being given by numbers 1, 19, 24, 6, 2, 20 in Table XXXIV, chosen at random. The heights of the contractions were as shown in Table XXXV.

TABLE XXXIV

The 24 possible arrangements of 4 letters

1 *ABCD*	7 *BDAC*	13 *ACDB*	19 *DCAB*
2 *BCDA*	8 *DACB*	14 *CDBA*	20 *CABD*
3 *CDAB*	9 *ADCB*	15 *DBAC*	21 *ADBC*
4 *DABC*	10 *DCBA*	16 *BACD*	22 *DBCA*
5 *ACBD*	11 *CBAD*	17 *ABDC*	23 *BCAD*
6 *CBDA*	12 *BADC*	18 *BDCA*	24 *CADB*

TABLE XXXV

Dose	*Heights of contraction in different groups*						Mean height
	1	2	3	4	5	6	
A	21	42	33	32	35	24	31
B	13	14	12	9	12	11	12
C	32	44	34	35	26	40	35
D	16	17	23	18	18	22	19

The calculation depends on the assumption that the height of contraction bears a linear relation to the log dose.

The mean of the two mean results obtained with the high doses *A* and *C* is the mean of 31 and 35, which is 33.

The mean of the two mean results obtained with the low doses *B* and *D* is the mean of 12 and 19, which is 15·5.

Then 33–15·5, which is 17·5, is the mean increase in height produced by multiplying the low dose by 4/3.

Next we obtain a measure of the difference in the effect of the standard and the unknown. The difference in the effect of *A* and *C* is 4 and the difference in the effect of *B* and *D* is 7, these differences being in the same direction.

Hence the mean difference between standard and unknown is 5·5.

Since the height of contraction is proportional to log dose, it follows that

$$\frac{17\cdot5}{5\cdot5} = \frac{\log 4/3}{M},$$

where *M* is the logarithm of the ratio of the potency of the two preparations in the doses injected.

But
$$\log 4/3 = 0\cdot6021 - 0\cdot4771$$
$$= 0\cdot125,$$

hence
$$M = \frac{0\cdot125 \times 5\cdot5}{17\cdot5}$$
$$= 0\cdot039$$

and antilog 0·039 is 1·09.

Thus since the unknown produced higher contractions than the standard

$$\frac{C}{A} = 1\cdot09,$$

or $$\frac{U}{10} = 1{\cdot}09 \times \frac{1{\cdot}5}{20} S,$$

therefore $$U = \frac{10{\cdot}9 \times 1{\cdot}5}{20} S = 0{\cdot}8175S.$$

But S is 0·5 unit per ml.; therefore U is 0·409 unit per ml.

This calculation is based on the work of Hemmingsen (1933). More elaborate methods are described in Chapter III and in Holton's paper (1948). The accuracy of the result can be determined as she explains, but a useful guide to the accuracy is given by the ratio of the two doses used. If the uterus is one in which a small increase in dose makes a large difference in the height of contraction, then the accuracy is correspondingly great; if a large increase in dose is necessary to obtain an appreciable increase in height, the accuracy is low.

Estimation on the fowl's blood-pressure

A method of estimating posterior lobe extract which has given good results in the hands of many workers has been described by Coon (1939). The property of causing a fall of blood-pressure in the fowl is used; this property is more strongly possessed by oxytocin than by vasopression.

Preparation of the bird. Coon used White Leghorn cockerels of weight from 1·8 to 2·2 kg. The cockerel is anaesthetized by injecting sodium phenobarbitone into the gastrocnemius muscle in a dose of 0·2 gm. per kg.; this injection is given 1–2 hours before the operation. When it is deeply anaesthetized, the cockerel is placed on its right side with the femur at right angles to the spine and the tibia, and its feet are secured. The feathers are removed from the outer surface of the left thigh, and a skin incision, 7–8 cm. long, is made 1·5 cm. below and parallel to the femur, and extending from a point about 2 cm. posterior to the distal end of the femur. This exposes the large gluteus primus muscle. The lower flap of skin is pulled down so that the posterior edge of the gluteus primus is seen lying over the semitendinosus muscle. This edge is separated from the underlying muscle for the length of the incision. When the free edge of the gluteus primus is lifted up the ischiadic artery and nerve and the crural vein are seen running along the anterior edge of the semitendinosus muscle. To make access easier the gluteus

primus is cut near the proximal end of the incision at right angles to its free edge. The corner of the resulting muscle flap is fastened to the skin of the upper thigh. Lengths of artery and vein are then prepared for inserting cannulae. As an anti-coagulant, 8·5 per cent. sodium citrate is recommended by Thompson (1944).

Performance of test. Injections are made at fixed intervals of 3 minutes. An extract containing 10 units per ml. is diluted 5 or 10 times, and of this dilution about 0·1 ml. is injected. The dose should be adjusted so as to produce a fall of blood-pressure of 20–40 mm. According to Coon's paper, the unknown solution is given alternately with the standard until a dose of the unknown is found which has the same effect as the standard. The method has been discussed and modified by Smith (1942), Smith and Vos (1943), and Thompson (1944), who has proposed an experimental design for the doses. An experimental design is certainly required and it seems likely that the design described above for the rat uterus method would be satisfactory, in which two doses of standard and two doses of unknown are chosen. Six groups of responses, each containing a response to each of these four doses, are then recorded. The calculation is the same as for the rat uterus.

II. Estimation of Antidiuretic Potency

For the estimation of the antidiuretic potency, dogs, rabbits, rats, mice, and leopard frogs (Boyd and Mack, 1940) have been used. Whatever animal is chosen the method must be comparative, for the reaction of any animal or group of animals to the administration of a given amount of the antidiuretic principle varies greatly at different times. A unit of activity defined as the amount of hormone required to delay the excretion of water in a dog or in mice for a certain length of time is therefore a variable and not a constant measure. Since the response of different animals differs, it is necessary to determine the average response in a group of animals, and smaller animals such as rats or mice are more convenient for this purpose than dogs or rabbits. Gibbs (1930) first proposed the use of mice. According to his method 8 mice are taken, and each is placed in a glass funnel on a floor of wide-meshed wire; each funnel is covered with a smaller inverted funnel. The funnels are placed in a convenient

rack and the urine is collected in graduated measures placed below. Each mouse receives an injection of 1 c.c. of warm tap-water into the peritoneal cavity. Four of the mice receive subcutaneous injections of pituitary extract, and 4 are kept as controls. After about 2 hours the control mice pass increasing quantities of urine, while the injected mice pass none. Gibbs proposed to discover the least amount of extract which delays the excretion of water for 3 hours.

The measurement of the volume of urine excreted by mice is relatively inaccurate since the amount is so small, and for

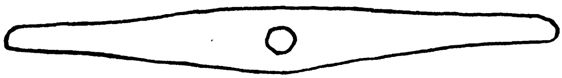


FIG. 24. Wooden gag for rat. 7 cm. long.

this reason rats are to be preferred. A method using rats has been described by Burn (1931).

Choice of animals. A group of 16 male rats is chosen of about the same weight; females are not taken because it has been found that their behaviour is more variable. The rats should not be less than 140 or more than 240 gm. Food is removed on the evening before a test is to be carried out, for if this is not done the results obtained are irregular.

Administration of water. When the test begins each rat is weighed and 5 c.c. water per 100 gm. is administered by stomach tube. A No. 3 gum elastic catheter is convenient for this purpose as this is too big to pass into the trachea. The catheter is attached by a piece of rubber tube to a 10-c.c. syringe filled with the water, which is previously warmed. The end of the catheter is dipped in glycerine before it is passed into the rat's stomach. A small wooden gag (see Fig. 24) with a central hole for the catheter is supported horizontally at one end by a clamp fixed on a stand. An assistant holds the rat vertically and passes the free end of the gag into the side of the rat's mouth. When the rat's mouth is opposite the hole, the free end of the gag is fixed in a second clamp. It is a help in passing the catheter if the rat's tongue is held forward between the gag and the lower jaw; this happens when the gag is put well back into the rat's mouth at the moment it is inserted. The catheter is then passed through the hole down the oesophagus, taking care

to use no force. If there is difficulty in passing the tube this is generally due to the skin at the back of the rat's neck being held too tightly by the assistant. A good hold of the rat is obtained by taking the skin at the back of the neck between the thumb and first finger and pressing the back of the second finger against the vertebral column of the rat.

Dosage of pituitary extract. So soon as the water has been given, pituitary extract is injected subcutaneously. An extract prepared from the standard powder containing 2 units per c.c. is diluted to contain 0·02 unit per c.c. Of this dilution 0·2 c.c. is injected per 100 gm. rat, so that the dose given is 0·004 unit per 100 gm. The sensitiveness of rats varies, and it may be found that this dose does not produce a sufficient delay in the diuresis. Usually, however, it will be enough. Of an unknown extract expected to contain 10 units per c.c., 0·2 c.c. is diluted to 100 c.c.; of this 0·2 c.c. is given per 100 gm.

Collection of urine. After injection of pituitary extract the rats are placed in circular cages (diameter 25 cm.) with a wire floor of large mesh, the cages resting on glass funnels which conduct the urine to a measuring cylinder below. Four rats are placed together in one cage, so that the rats injected with the standard occupy two cages, and those injected with the unknown occupy two more. The urine is collected in 10 c.c. measures which are emptied when necessary into larger measures.

Estimation of the antidiuretic effect. The time taken for administration of water and injection of pituitary extract to a group of 4 rats is usually about 6 to 8 minutes. The mid-point of this period is noted and taken as the starting-point for the group. The method of calculating the interval which elapses until the point of maximum rate of excretion of the water will be best understood by considering the examples in Table XXXVI. The rats in Group I received water and were injected at 11 a.m. An excretion of urine was first noticed at 12.30 p.m., when 1·4 c.c. was recorded; thereafter the total volume excreted was recorded every 15 minutes until 3 p.m., when the excretion for the last four periods had been consistently small, and it was concluded that further observations were unnecessary. In the third column of Table XXXVI is shown the amount of urine excreted per 15 minutes. These amounts, when added together, come to 22·9 c.c.; this figure excludes the amount excreted first

of all. The point of maximum rate of excretion was taken to be the point at which half of this quantity of 22·9 c.c. was excreted. The half of 22·9 is 11·4; to this was added the amount of 1·4 c.c. originally excreted, making in all 12·8 c.c. This total excretion was reached between 1.15 and 1.30 p.m. At 1.15 p.m. the total excretion was 7·8 c.c., and at 1.30 p.m. it was 14·8 c.c.; the excretion of 12·8 c.c. was reached at a point between 1.15 and 1.30 p.m. dividing the 15-minute period in the ratio of 5 to 2, that is to say, at approximately 1.25 p.m. From 11 a.m. to 1.25 p.m. is 145 minutes, and this time in minutes was taken as the time to maximum rate of excretion. In this particular experiment it is clear that the point of maximum rate of excretion could have been picked out at once by inspection of the figures of the third column of Table XXXVI; in other experiments, however, such as that recorded for the second group of rats in the same table, inspection is little help, and a fixed rule for discovering the point of maximum rate of excretion is necessary.

It may be asked why the point at which half the total excretion is reached is not taken as the point of maximum rate of excretion, to which the answer is that occasionally an initial amount of 1·0 to 3·0 c.c. of urine may be excreted, followed by an interval of ½ to 1 hour in which there is no further excretion. This initial amount does not appear to belong to the true period of excretion, and it should not be allowed to complicate the determination of the mid-point of that period. Admittedly, however, in the majority of experiments the point of maximum rate of excretion is close to the point at which half the total volume is excreted. Further, it may be asked whether the length of time for which the excretion is recorded does not affect the determination of the point of maximum rate of excretion; provided it is evident that the excretion is nearly complete, so that, for a group of 4 rats, a total of 20 to 30 c.c. of urine has been obtained, then it is practically immaterial how long the observation is continued, and the interval to the maximum rate of excretion will not vary by more than a minute or two.

Calculation of the result. If the unknown extract is compared with the standard only once, the comparison is subject to the error introduced by the difference in sensitiveness of the two groups of animals. The comparison is, therefore, made again,

TABLE XXXVI

First group of 4 rats			*Second group of 4 rats*		
Water administered and pituitary extract injected at 11 a.m.			*Water administered and pituitary extract injected at 11.25 a.m.*		
Time	*Total urine*	*Volume in last 15 minutes*	*Time*	*Total urine*	*Volume in last 15 minutes*
12.30	1·4	. .	12.05	0·5	. .
12.45	1·6	0·2	12.20	0·5	. .
1.00	4·2	2·6	12.35	0·5	. .
1.15	7·8	3·6	12.50	2·2	1·7
1.30	14·8	7·0	1.05	3·5	1·3
1.45	17·0	2·2	1.20	6·3	2·8
2.00	20·5	3·5	1.35	8·2	1·9
2.15	21·6	1·1	1.50	13·0	4·8
2.30	23·2	1·6	2.05	14·4	1·4
2.45	24·3	1·1	2.20	17·5	3·1
3.00	24·3	0·0	2.35	20·0	2·5
. .	. .	. .	2.50	22·0	2·0
. .	. .	. .	3.05	22·0	0·0
Time to maximum rate of excretion 145 minutes			Time to maximum rate of excretion 140 minutes		

2 or 3 days later, when the rats which previously were injected with the standard are injected with the unknown, and those previously injected with the unknown are injected with the standard. The figures in Table XXXVII furnish an example of the results obtained.

TABLE XXXVII

Date	*Groups of rats*	*Extract injected*	*Time to maximum excretion*
23 Sept.	A	Standard	136 minutes
	B	,,	131 ,,
	C	Unknown	176 ,,
	D	,,	144 ,,
25 Sept.	A	Unknown	121 ,,
	B	,,	124 ,,
	C	Standard	166 ,,
	D	,,	149 ,,

The average figure for all rats injected with the standard on the 2 days is 145 minutes, while the average figure for rats injected with the unknown is 141 minutes. The relative potency of the doses producing these effects is determined from the

graph shown in Fig. 25 which relates the log dose to the time to maximum excretion. The graph shows that the ordinate 145 corresponds to an abscissa of $\bar{3}$·648 and the ordinate 141 to $\bar{3}$·593. Now $\bar{3}$·648 is the log. of 0·0044, and $\bar{3}$·593 is the log. of 0·0039.

The dose of unknown injected was 0·0006 c.c. per 100 gm.,

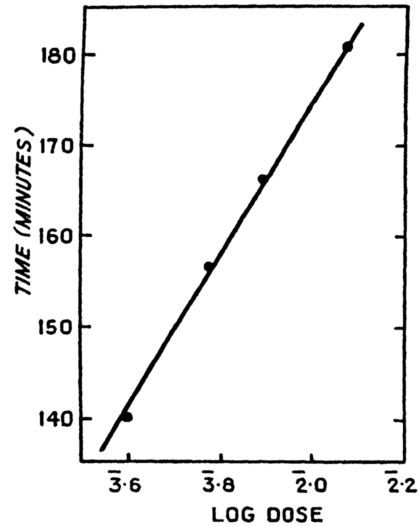


FIG. 25. Assay of antidiuretic hormone. Graph showing relation of log dose (unit per 100 gm. body-weight) to time of maximal rate of excretion. (Burn.)

while the dose of standard corresponded to 0·003 mg. standard powder per 100 gm. Hence

$$\frac{0\cdot0006 \text{ c.c. unknown}}{0\cdot003 \text{ mg. standard}} = \frac{0\cdot0039}{0\cdot0044}$$

or

$$1 \text{ c.c. unknown} = 4\cdot5 \text{ mg. standard}$$
$$= 9 \text{ units.}$$

Curve relating dose to effect. The curve given in Fig. 25 was constructed from observations on 64 male rats which were made on 4 days. On each day the doses 0·004, 0·006, 0·008, and 0·012 unit per 100 gm. were each administered to a group of 4 rats. The results are given in Table XXXVIII.

TABLE XXXVIII

Dose of pituitary extract, unit per 100 gm.	*Different effects in different groups of 4 rats (minutes)*	*Average effect (minutes)*
0·004	126, 138, 144, 152	140
0·006	140, 145, 178, 163	156·5
0·008	191, 156, 159, 159	166
0·012	210, 174, 163, 175	180·5

The log doses are plotted against the average effect in Fig. 25. Those who propose to use the method extensively should determine the curve in their own laboratory.

Error of the method. Burn (1931) has given figures for eight comparisons where the true relative results were known; the estimated results varied from the true results by amounts up to 23 per cent., the average variation being 12 per cent.

Ham (1943) plots the excretion of water at 15-minute intervals for 5 hours on a chart and obtains the area of the excretion curve; he also draws attention to the need that rats should be fully hydrated at the beginning of experiments.

REFERENCES

BEARVILLAIN, 1943, *Comp. Rend. Soc. de Biol.* **137**, 284.
BOYD and MACK, 1940, *Endocrin.* **26**, 153.
BURN, 1931, *Quart. J. Pharm. Pharmacol.* **4**, 517.
COON, 1939, *Arch. Internat. de Pharmacodyn. et de Thérap.* **62**, 79.
DALE and LAIDLAW, 1912, *J. Pharmacol. Exp. Ther.* **4**, 75.
GADDUM, 1927, *Pharmaceut. J.* **119**, 580.
GIBBS, 1930, *J. Pharmacol. Exp. Ther.* **40**, 129.
HAM, 1943, *Proc. Soc. Exper. Biol. and Med.* **53**, 210.
HAMBURGER, 1945, *Acta Pharmacol. and Toxicol.* **1**, 112.
HEMMINGSEN, 1933, *Quart. J. Pharm. Pharmacol.* **6**, 187.
HOLTON, 1948, *Brit. J. Pharmacol.* **3**, 328.
DE JALON, BAYO, and DE JALON, 1945, *Farmacoterap. Actual,* **2**, 313.
MORRELL, ALLMARK, and BACHINSKI, 1940, *J. Pharmacol. Exp. Ther.* **70**, 440.
—— —— —— 1945, *Canad. J. Research*, Sect. E, **23**, 126.
SCHILD, 1947, *Brit. J. Pharmacol.* **2**, 189.
SMITH, 1942, *J. Pharmacol. Exp. Ther.* **75**, 342.
SMITH and MCCLOSKY, 1924, *Hyg. Lab. Bull.*, No. 138.
SMITH and VOS, 1943, *J. Pharmacol. Exp. Ther.* **78**, 72.
THOMPSON, 1944, ibid. **80**, 373.

CHAPTER V

INSULIN

THE standardization of insulin has played a large part in the development of biological methods. The clinical importance of accurate standardization of insulin is greater than that of any other substance, for a difference of 10 per cent. in the potency of insulin has serious consequences for diabetic patients. If the potency is higher than usual some of them will suffer from hypoglycaemic reactions, while if it is less they will lose more dextrose in the urine. As a result the standardization of insulin has been studied in great detail, with results which can only be given briefly in this chapter. Those who work regularly with insulin should consult the papers of Hemmingsen and Marks (1932), Marks (1932), Hemmingsen (1933), Hemmingsen and Marks (1933), and also the *Quarterly Bulletin of the Health Organization*, League of Nations, Special Number, November 1936. This contains details of the methods employed in six different laboratories to compare the international standard (1935) with the previous international standard (1925).

The standard preparation. The first international standard for insulin was adopted as the result of a conference which met in Geneva in 1925; it was a dry stable preparation of which 0·125 mg. contained, by definition, 1 unit of activity. The present international standard was adopted in 1935. It is a sample of the purest insulin obtainable by recrystallization of the crystalline substance. The whole of the material was prepared by the Insulin Committee of the University of Toronto in the Connaught Laboratories. The discovery of D. A. Scott that crystalline insulin is a zinc salt made it possible to provide about 51 gm. of this pure material.

For the North American Continent the standard is kept for distribution by the Insulin Committee of the University of Toronto; for the rest of the world it is kept for distribution in the National Institute for Medical Research, London.

The standard is stored in tubes each of which contains approximately 20 mg. In these tubes it was kept for some weeks *in vacuo* over phosphorus pentoxide until completely dry, when

the tubes were filled with dry nitrogen and sealed. The storage is below 0° C. and is in the dark.

The unit. As a result of a comparison of this new standard with the former standard in different laboratories, it was decided that the unit should be defined as 'the specific insulin activity of one twenty-second (1/22) of a milligramme of the new Standard (1935)'. Hence the present international standard contains 22 units per mg.

The use of the standard. To prepare a solution of the standard, an ampoule is opened with a file and the contents transferred to a stoppered weighing-bottle for weighing. The weighed standard insulin is then dissolved in distilled water which contains 0·85 per cent. sodium chloride, sufficient hydrochloric acid to make pH 2·5, and 0·3 per cent. tricresol to prevent the growth of moulds. The volume of solvent should be enough to give, for example, 20 units per c.c. If this solution of the standard is kept in a stoppered vessel at the lowest temperature which will not freeze it, it is stable for at least a year.

Methods of Comparison

There are two methods of comparing samples of insulin, the rabbit blood-sugar method and the mouse-convulsion method. These methods usually give the same result except when an impure preparation of insulin is compared with one which is highly purified. For example, when the crystalline insulin which is now the international standard (1935) was compared with the former international standard, it was observed by several workers that by the rabbit method the crystalline insulin contained 20–22 units per mg., whereas by the mouse method it contained 24–28 units per mg. The figures obtained in different laboratories are given in Table XXXIX. Insulin as produced commercially to-day is highly purified, and therefore, when compared with the new international standard (1935), will usually give the same result by both methods.

As time has passed the mouse-convulsion method has superseded the rabbit blood-sugar method partly on account of its greater simplicity, and partly on account of the requirement in Great Britain by the Ministry of Health under the Therapeutic Substances Act. The assay must be conducted so that the 95 per cent. fiducial limits of error of the estimated potency may be

calculated from the figures obtained in that assay; the lower of these 95 per cent. fiducial limits of error must be greater than 80 per cent. of the stated potency of the batch and the upper must be less than 125 per cent. of the stated potency. The mean must be within ±10 per cent. of the stated potency. This requirement can be met most economically by using the mouse-convulsion method.

A. The Rabbit Test

This test was first introduced by Marks (1926); it depends on the observation of the fall of blood-sugar produced by insulin.

TABLE XXXIX

Taken from the *Quart. Bull. Health Organization* L. N. Special Number, Nov. 1936

Investigating Laboratory	*Rabbit method* *Result Units per mg.*	*Rabbit method* *Diluent*	*Mouse method* *Result Units per mg.*	*Mouse method* *Diluent*	*Temperature of incubator in mouse test*
Toronto (Hershey and Lacey)	21·3	Acid saline	24·8	Acid saline	37·5°
(Scott) .	..	..	24·0	Acid saline	37·5°
'A.B.' Insulin Laboratories	21·9	Acid saline	28·0	Acid saline	37°
Wellcome Laboratories	..	..	27·04	Neutral saline	34°
Copenhagen .	..	..	25·4	Acid saline	29°
Indianapolis .	21·9	Acid saline	..	..	..
Hampstead .	20·4	Water	20·6	Water	29°
Weighted mean	21·5	..	Excluding Hampstead 25·2	..	..

Selection and preparation of animals. Rabbits should be chosen 1·5–2·0 kg. in weight, with fairly large ears. They should be given a diet of oats (about 100 gm. daily), hay, and green-stuffs. For each test 12–20 rabbits are required, and it is important in the first place to discover whether any of them readily go into convulsions. They are prepared for a test by taking

away food from their cages in the afternoon of one day (water is left in the cages), and on the next day each rabbit is given a subcutaneous injection of 1 unit of insulin; it is then observed during 5 hours to see if it passes into convulsions. About 15 per cent. of rabbits have convulsions when injected with this dose of insulin, and before the insulin comparison is made, convulsing rabbits must be replaced by others which do not have convulsions.

Performance of test. On another day when the rabbits have again been prepared by withholding food overnight, they are divided into two equal groups. Each of the first group receives an injection of 1 unit of the standard powder dissolved and diluted so that the volume injected is about 0·5 c.c. The injection is made under the loose skin of the flank or back. During the period in which the test is completed, the animals' weights do not change, so it is unnecessary to adjust the dose according to the weight of the animal. Each of the second group receives such a dose of the unknown as may be expected to contain 1 unit of activity.

Samples of blood from each rabbit are obtained before the injections of insulin are made, and at hourly intervals for 5 hours after the injection. It is of great importance that the dose of insulin given should not be more than 0·5 unit per kg., that is, a dose of 1 unit for a rabbit of 2·0 kg. weight. Marks has determined the relation between dosage and hypoglycaemic effect, and has found (see Fig. 26) that when doses above 0·75 unit per kg. are given there is no further increase of effect. According to Fig. 26 higher doses must produce effects on the horizontal part of the curve, which are not proportional to the doses, and which cannot give results of value in determining the potency of an unknown sample.

Collection of blood samples. To collect samples of blood from a rabbit, it is not necessary to secure the rabbit; some workers, however, place it on a board between two upright pieces of wood, so that the rabbit cannot move sideways; others put it in a box with a hole at one end through which its head protrudes. One ear is held forward from the rabbit, and the hair on the outer edge is lightly shaved off with a sharp razor. The ear should then be thoroughly warmed. This is best done by placing an electric lamp close to the ear for a minute or two, or by rubbing.

The vein running along the outer edge should then stand out and remain distended for a few minutes.

It should be emphasized that the secret of obtaining blood easily depends on the thoroughness of the preliminary warming, and it is wise to spend time on this before puncturing the vein. A small artery-clip is placed on the vein near the base of the

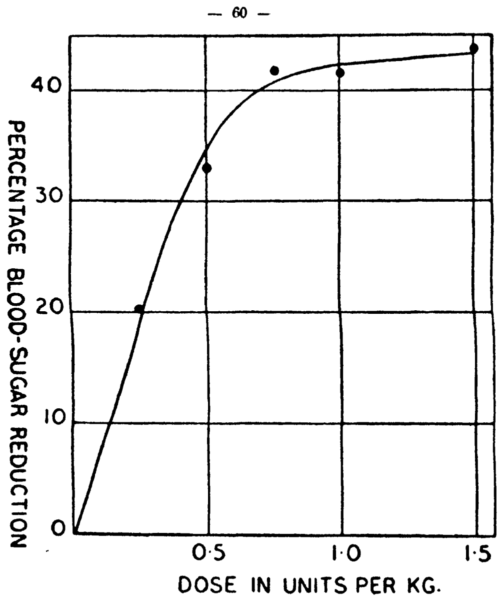


FIG. 26. Curve obtained on 40 rabbits relating the dose of insulin to the percentage blood-sugar reduction. If the dose exceeds 0·75 unit per kg. there is no further increase in the effect on the blood-sugar. (Marks, *League of Nations Report on Insulin Standardization*, C.H. 398.)

ear, and the vein is punctured either with a sharp triangular needle or with a scalpel of the pattern used in ophthalmic surgery. The puncture is made along the line of the vein towards the base of the ear, the needle entering at a point 1–2 cm. from the artery-clip. A rapid succession of drops should issue from the puncture when the needle is withdrawn, 25 to 30 drops usually giving 1 c.c. of blood. Sometimes, after an initial rapid flow, the drops come very slowly or may cease; the flow may then be re-established by rubbing the middle of the ear near the base with one finger. If a clot forms at the puncture, it

should be wiped off with a piece of cotton-wool, and the needle inserted to open the puncture again. In order to preserve the vein, as little use as possible of the needle should be made. When sufficient blood is collected, the artery-clip is at once removed, and a piece of cotton-wool is held on the puncture until the bleeding stops.

The blood is collected in glass tubes about 3·5 cm. high and 2·5 cm. diameter with a flat base; potassium oxalate is put in each tube beforehand, taking about 5 mg. per c.c. blood.

Determination of blood-sugar. The procedure used by Marks for determining blood-sugar is a combination of Somogyi's (1930) method of precipitating the proteins with Hagedorn and Jensen's (1923) method of estimating sugar.

Solution A.

Zinc sulphate ($ZnSO_4$, $7H_2O$)	25 gm.
N. sulphuric acid	62·5 c.c.
Distilled water	to 2,000 c.c.

Solution B.

N. sodium hydroxide	100 c.c.
Distilled water	to 2,000 c.c.

This solution must be adjusted so that 8 c.c. exactly neutralizes 4 c.c. solution *A*.

When mixed in these proportions the mixture should not be red to phenolphthalein, and the clear supernatant fluid should not give a further precipitate of zinc hydroxide on the addition of more of solution *B*.

Solution C.

Potassium ferricyanide	0·825 gm.
Anhydrous sodium carbonate	5·3 gm.

Dissolved in distilled water and diluted to 1,250 c.c.

Solution D.

Acetic acid (glacial)	3 c.c.
Distilled water	to 100 c.c.

Solution E.

10 c.c. of solution E_1 to 90 c.c. of solution E_2.

Solution E_1.

Potassium iodide .	25 gm.
Distilled water	to 100 c.c.

Solution E_2.

Zinc sulphate .	25 gm.
Sodium chloride .	125 gm.
Distilled water .	to 450 c.c.

Solution F.

N/10 sodium thiosulphate	5 c.c.
Sodium bicarbonate	0·01 gm.
Distilled water	to 100 c.c.

Starch indicator is prepared by making 1 gm. soluble starch into a paste with a little distilled water and pouring into 100 c.c. boiling saturated NaCl.

Procedure. 4 c.c. solution *A* is put in a test-tube and 0·5 c.c. blood is delivered into it. 8 c.c. of solution *B* is added, the test-tube is shaken and placed in a boiling water-bath for 2–3 minutes. The liquid is filtered through a dry paper. Take 5 c.c. filtrate and 5 c.c. solution *C* in a boiling-tube with a small glass funnel resting on the top to minimize evaporation. Place in a boiling water-bath for 15 minutes. Cool, add 2 c.c. solution *D* and 3 c.c. solution *E*. Titrate with solution *F* from a burette, using the starch as indicator.

The blank titration, when this procedure is followed without the addition of blood filtrate, i.e. in the absence of sugar, should be precisely 2 c.c. Table XL gives the percentage of sugar in the blood corresponding to amounts of thiosulphate used less than 2 c.c. This table, however, is prepared for the ordinary Hagedorn-Jensen method in which only 0·1 c.c. of blood is used. Since 5 c.c. of the blood filtrate are added, corresponding to 0·2 c.c. of blood (since 12·5 c.c. contain 0·5 c.c. blood, 5 c.c. contain 0·2 c.c.), the figure for the blood-sugar percentage obtained from the table must be divided by two.

TABLE XL

c.c. N/200 thiosulphate	*Percentage blood-sugar*	*c.c. N/200 thiosulphate*	*Percentage blood-sugar*
0·0	0·385	1·0	0·177
0·1	0·355	1·1	0·159
0·2	0·331	1·2	0·141
0·3	0·310	1·3	0·124
0·4	0·290	1·4	0·106
0·5	0·270	1·5	0·088
0·6	0·251	1·6	0·070
0·7	0·232	1·7	0·052
0·8	0·213	1·8	0·034
0·9	0·195	1·9	0·017

The intermediate values are obtained by plotting the above

figures on a graph, using c.c. N/200 thiosulphate as abscissae, and percentages of blood-sugar as ordinates.

Pooling of samples. It is unnecessary to determine the percentage of the blood-sugar in each of the five samples collected at hourly intervals after the insulin is injected, and it is sufficient to prepare two duplicate pooled samples in the following way. At each bleeding rather more than 0·2 c.c. is collected in tubes containing potassium oxalate, and of this 0·1 c.c. is put into each of two test-tubes containing 4 c.c. of solution *A*. The same two test-tubes are used for all five bleedings from any one rabbit, so that at the end of the experiment each test-tube contains 0·5 c.c. of blood containing a percentage of sugar which is the average of that present in the five separate samples. The proteins are precipitated by the addition of 8 c.c. solution *B* as before.

The cross-over test. On the first day of the test half of the rabbits are injected with 1 unit of standard insulin. and half are injected with an amount of the unknown sample expected to contain 1 unit. On the second day of the test the groups are crossed over, so that those which received the standard now receive the unknown, and those which received the unknown now receive the standard. If the second day of the test is actually the day following the first day, the rabbits are then kept without food between the two parts of the test. There can, however, be an interval between the two days on which the test is performed.

Expression of the hypoglycaemic reaction. The method of calculating the hypoglycaemic effect in a given rabbit is as follows. If the initial blood-sugar was 0·106 per cent. and the average of the five bleedings was 0·080 per cent., then the average fall during 5 hours is 0·026 per cent. This is expressed as a percentage of the initial value to give a figure which is called the percentage blood-sugar reduction, which is, in this example

$$\frac{0{\cdot}026}{0{\cdot}106} \times 100 = 24{\cdot}5 \text{ per cent.}$$

Table XLI is an example of a complete assay carried out on 12 rabbits. In the first experiment each rabbit receiving the standard was injected with 1 unit of insulin, and each rabbit receiving the unknown sample was injected with an amount expected to contain 1 unit. In this experiment the responses

Table XLI. *Experiment 1*

1st day.

Rabbits on standard. Dose 1 unit per rabbit.

Rabbit no.	*Weight of rabbit gm.*	*Initial blood-sugar mg. per 100 c.c.*	*5-hour pool blood-sugar mg. per 100 c.c.*	*Percentage blood-sugar reduction*
1	1,800	112	60	46·4
2	2,000	106	54	49·0
3	2,150	102	73	28·4
4	1,950	102	60	41·1
5	1,800	106	66	37·7
6	1,900	106	54	49·0
				251·6

Rabbits on test. Dose 0·5 c.c. per rabbit of sample diluted 20 times.

7	1,750	112	70	37·5
8	2,000	110	64	41·8
9	1,900	102	54	47·1
10	1,900	116	76	34·5
11	1,550	106	66	37·7
12	2,500	100	50	50·0
				248·6

2nd day.

Rabbits on test. Dose 0·5 c.c. per rabbit of sample diluted 20 times.

1	1,800	86	40	53·5
2	2,000	90	40	55·5
3	2,150	100	66	34·0
4	1,950	100	66	34·0
5	1,800	96	70	27·1
6	1,900	104	54	48·1
				252·2

Rabbits on standard. Dose 1 unit per rabbit.

7	1,750	100	66	34·0
8	2,000	94	66	29·8
9	1,900	104	46	55·8
10	1,900	112	66	41·1
11	1,550	96	46	52·1
12	2,500	94	64	31·9
				244·7

Total for rabbits receiving standard . . . . 496·3
Total for rabbits receiving test . . . . . 500·8
Hence the effect of the test is 101 per cent. of the effect of the standard.

Experiment 2

1st day.

Rabbits on standard. Dose 0·5 unit per rabbit.

Rabbit no.	*Weight of rabbit gm.*	*Initial blood-sugar mg. per 100 c.c.*	*5-hour pool blood-sugar mg. per 100 c.c.*	*Percentage blood-sugar reduction*
1*a*	2,000	110	92	16·4
2	2,250	92	64	30·4
3	2,250	112	80	28·6
4	1,850	100	66	34·0
5	1,950	92	66	28·2
				137·6

Rabbits on test. Dose 0·5 c.c. per rabbit of sample diluted 40 times.

7*a*	2,300	86	54	37·2
8	2,200	100	76	24·0
9*a*	1,950	102	80	21·5
10	2,000	106	80	24·5
12	2,250	92	52	43·5
				150·7

2nd day.

Rabbits on test. Dose 0·5 c.c. per rabbit of sample diluted 40 times.

1*a*	2,000	110	80	27·3
2	2,250	86	60	30·2
3	2,250	100	72	28·0
4	1,850	110	86	21·9
5	1,950	100	80	20·0
				127·4

Rabbits on standard. Dose 0·5 unit per rabbit.

7*a*	2,300	96	54	43·8
8	2,200	100	76	24·0
9*a*	1,950	96	80	16·6
10	2,000	94	72	23·4
12	2,250	86	46	46·5
				154·3

Total for rabbits receiving standard . . . . 291·9
Total for rabbits receiving test . . . . . 278·1
Hence the effect of the test is 95·2 per cent. of the effect of the standard.

were high, that is to say, the average percentage reduction of blood-sugar was about 40. A second experiment was therefore performed in which each rabbit was injected with 0·5 unit of standard and an amount of the unknown sample expected to contain 0·5 unit. The average percentage blood-sugar reduction was 25–30.

Calculation of the result. The results obtained in these two experiments can be used to illustrate several points. The first experiment indicated that the unknown and the standard were equal in the doses injected, while the second indicated that the unknown was weaker. This is because the doses given in the first experiment were too high, and the effects were nearly maximal. Yet these doses were only 1 unit per rabbit, or approximately 0·5 unit per kg. In the second experiment, in which the percentage reductions were 25–30, the doses were 0·5 unit per rabbit. Now in the Insulin Committee Laboratory in Toronto the dose ordinarily given is 2 units per rabbit of 2 kg.; it must be concluded that unless rabbits in Toronto are much more resistant to insulin than rabbits in London, the dose used in Toronto is too high.

(*a*) *Correction for the initial blood-sugar.* The results can be corrected for differences in the initial blood-sugar. In Experiment 2 the sum of the initial blood-sugars in rabbits injected with the standard was 978 (mg. per 100 c.c. blood); the sum in rabbits injected with the unknown sample was 992. That is to say, the rabbits injected with the unknown sample started with higher initial blood-sugars. Now if the figure for percentage reduction were an ideal measure of the insulin effect, it would be independent of the figure for the initial blood-sugar. In fact it is not independent, since Hemmingsen and Marks (1932) have shown that the figure for the percentage blood-sugar reduction increases with increasing initial value of the blood-sugar. Hence, if the initial blood-sugar values are higher for the unknown sample than for the standard, the percentage reduction figures for the sample will be relatively too high. Some workers neglect the difference in the initial values, since there is no way of making an accurate correction. The best approach to a correction is to find the difference between the total figures for the initial sugar values, to multiply this difference by 0·22, and to add the result to the total of the percentage reductions on the

side showing the lower total value of initial blood-sugar. Thus in Experiment 2 the difference between the total initial figures is 992−978 = 14. When 14 is multiplied by 0·22 the result is 3·08. This figure is added to the total percentage reduction for the standard, since the standard has the lowest total initial sugar value. Hence the percentage reduction for the standard

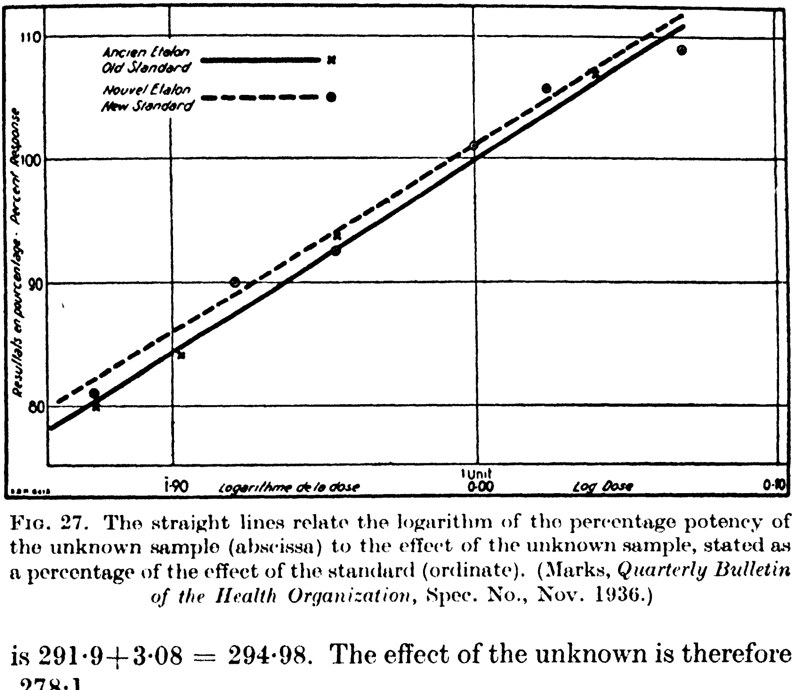


FIG. 27. The straight lines relate the logarithm of the percentage potency of the unknown sample (abscissa) to the effect of the unknown sample, stated as a percentage of the effect of the standard (ordinate). (Marks, *Quarterly Bulletin of the Health Organization*, Spec. No., Nov. 1936.)

is 291·9+3·08 = 294·98. The effect of the unknown is therefore $\frac{278 \cdot 1}{294 \cdot 98} \times 100$ of the standard, or 94 per cent. of that of the standard.

(*b*) *Relation between dose and effect*. Having obtained the corrected figure for the relation of the effect of the unknown sample to the effect of the standard, the worker must then determine the relative potency of the unknown sample and the standard. The evidence is strong that this relation is not linear as has been assumed by some. Marks has twice determined the relation at an interval of 10 years on preparations differing widely in purity, and since his results were the same each time, they can be accepted as correct. In Fig. 27 is reproduced a graph from Marks's paper in the *Quarterly Bulletin of the Health Organization*,

The source of error is well illustrated by the results of Walden in the *Quarterly Bulletin* (1936). In testing crystalline insulin against the former international standard, he first assumed that the crystals contained 24 units per mg.; he found 22·9 units per mg. He then assumed 22 units per mg., and found 21·9. Finally, he assumed 20 units per mg. and found 20·9. Thus in these results, which were obtained on 113, 55, and 119 rabbits respectively, the result depended on the assumption made. This dependence of result on assumption, which makes the method appear ridiculous, would have been very much smaller if the correct relation shown in Fig. 27 had been employed.

Continuous testing of insulin. Workers who are constantly testing insulin can determine the relation shown in Fig. 27 for rabbits in their own laboratory in the course of ordinary tests. They can do this by using two doses of the standard, namely, 0·9 unit and 1·1 unit per rabbit, instead of 1 unit. Half the rabbits which receive the standard receive the lower dose and half the higher dose. This difference of dose can be neglected in calculating the potency of the unknown sample, but as tests are completed, data will be accumulated for determining the position of the two points 0·9 and 1·1 in Fig. 27. A straight line through these points will give the relation of effect to potency.

B. The Mouse Method

The mouse-convulsion method is in principle very simple; it consists in finding a dose of the unknown sample which produces convulsions in the same proportion of mice as a dose of the standard. The account which is given is based on information obtained from Trevan and Marks and from the papers of Hemmingsen.

Conditions of the experiment. The two conditions which are important in carrying out the test are the temperature and the period in which the mice are kept without food. At ordinary room temperatures, even when a large dose of insulin is injected, a proportion of mice never convulse. To enable the convulsion rate to approach 100 per cent., a raised temperature is necessary. Trevan works at 34° C., while Hemmingsen works at 29° C. Similarly, the period of fasting plays a part; Trevan uses mice which have been kept without food for 17–22 hours; Hemmingsen uses mice kept without food for 1½ hours. Hemmingsen and

Marks (1933) state that the slope of the line relating the dose of insulin to the percentage of mice convulsing is steeper for mice kept without food for 24 hours and tested at 37° C. than it is for mice kept without food for 1½ hours and tested at 29° C. If the period of fasting is the same, they state that the difference of temperature alone does not affect the slope of the line.

If experiments are performed under Trevan's conditions, more mice die and fewer are available for subsequent experiments. Hence, if the supply of mice is restricted, it is better to work under Hemmingsen's conditions. On the other hand, it can be argued that repeated use of the same mice is on general grounds undesirable, and, if the supply of mice is short, that it is better to use the rabbit method.

Apparatus for temperature control. Different forms of apparatus for keeping the mice at the desired temperature are used. Trevan described a water-bath, and Hemmingsen an air-bath. The water-bath is the simplest, and is therefore given first.

The mice are put in zinc boxes about 12·5 cm. cube, which are for three-quarters of their height immersed in a water-bath at 34° C. The boxes are covered with glass lids and perforations are made around the top of the walls for ventilation (Fig. 28). One water-bath holds 10 boxes, each of which takes 4 mice. The water in the bath is kept at 34° C. by 4 carbon-filament heating lamps and constantly stirred by air blown in through a pipe as shown in the figure. The apparatus is protected from draughts.

The air-bath used by Marks is based on that devised by Krogh and Hemmingsen; it consists of a wooden cupboard filled with shelves on which are placed glass jars containing the mice. The mice are observed through glass in the doors. The cupboard has compartments on each side for the mice, with a central portion without shelves in which there is an electric fan for producing a circulation of air, and a thermo-regulator. The inside measurements of the cupboard are: height, 114 cm., width 147 cm., while the depth from back to front is 14 cm. The central portion where there are no shelves is 15 cm. wide. The framework is constructed of wood 1·6 cm. thick, and the cupboard, except in front, is encased in non-conducting material. The front is made of 4 doors with glass panels and 1 wooden door. The doors with glass are upper and lower doors to each side compartment, and the wooden door is to the central air-chamber. There are 7 shelves on each side, so that with 6 jars on each shelf, and 2 mice in each jar, the total capacity is 168 mice. Normally, the bottom shelf is not used. The shelves are strips of wood 4·3 × 1·6 cm. in cross-section, separated by a space of 5 cm.,

bridged underneath by perforated zinc. This construction allows free circulation of air through the cupboard, provides ample ventilation for the mice, but does not allow them to escape from the jars, which must just fit between the shelves.

The cupboard is heated by resistance wire running in glass tubing

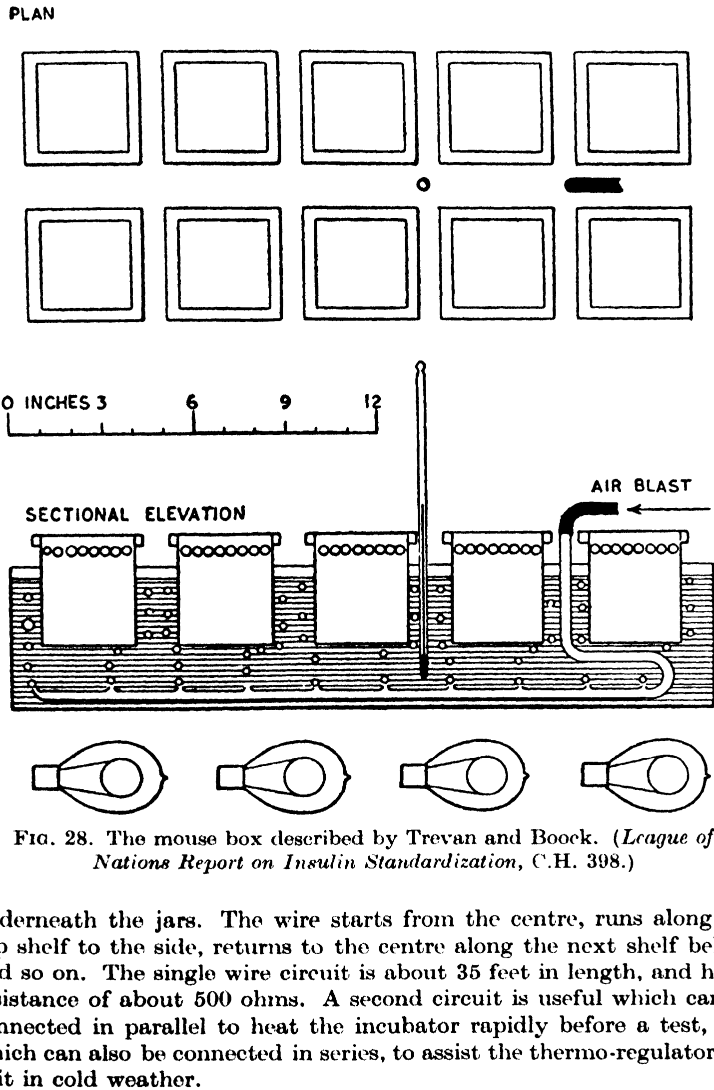


FIG. 28. The mouse box described by Trevan and Boock. (*League of Nations Report on Insulin Standardization*, C.H. 398.)

underneath the jars. The wire starts from the centre, runs along the top shelf to the side, returns to the centre along the next shelf below, and so on. The single wire circuit is about 35 feet in length, and has a resistance of about 500 ohms. A second circuit is useful which can be connected in parallel to heat the incubator rapidly before a test, and which can also be connected in series, to assist the thermo-regulator circuit in cold weather.

The fan-motor is supported on tightly stretched rubber pressure tubing, so that vibration in the incubator is eliminated.

Preparation of the mice. Mice are fed on bread and milk, or bread, for a few days before the test. The effect of diet on the response of mice to insulin was first studied by Bainbridge (1925) and has been more recently described by Rowlinson and Lesford (1948), who have shown that the accuracy obtained with a given number of mice is greater on a bread diet than on a diet consisting of cubes of material containing more protein and more fat. It is probable that the difference does not only depend on the composition of the diet but also on the greater ease with which the mice can eat and digest cooked food like bread. On a bread diet the weight of a result, depending on the steepness of the slope of the line relating dose to percentage convulsions, is much greater. Food is then removed from the cages either at 5 p.m. on the day before the test or 1½ hours before the test begins, according to the time preferred. The mice should be weighed in order that the two groups into which they are divided should contain equal numbers of mice of any given weight. In the comparison of crystalline insulin with the former standard, the dose used by Trevan was 0·4–0·5 μg. crystalline insulin, or 0·009–0·011 unit injected in a volume of 0·5 c.c. saline per 20-gm. mouse. The dose used by Hemmingsen was similar; he was careful to use mice of similar weight in any one test. Marks used two doses, the larger twice as great as the smaller; the larger dose used was from 0·023 to 0·03 unit per mouse given irrespective of body-weight. His doses were larger because his injections were made in distilled water. In ordinary tests all dilutions should be made with saline acidified to pH 4.

Method of making the test. The following example is taken from Rowlinson and Lesford, being based on the work of Marks and Pak (1936). Mice were kept without food from 5 p.m. of the previous day. Ninety-six mice between 15 and 23 gm. were used and divided into 4 groups of 24, taking care to see that the distribution of mice of different weights was the same in each group. Standard insulin solutions were prepared containing 48 and 80 milliunits per c.c., giving a dose ratio of 1·2 to 2. The sample under test was similarly diluted, assuming a value for the potency. The same volume of insulin was given to all mice in any test, and was within the range of 0·25 to 0·4 c.c. per mouse. The following results were obtained in one test.

High dose of standard	28 m U per mouse.	
Low ,, ,,	16·8 ,, ,,	

Response

Standard high dose	17 out of 24	
,, low ,,	4 ,, 24	
Test high dose	19 ,, 24	
,, low ,,	8 ,, 24	

The calculation of the result is based on the work of Hemmingsen (1932), who showed that the relation between the logarithm of the dose and the percentage of convulsions corresponds to a normal integrated frequency curve. Between 15 and 85 per cent. this frequency curve is very nearly a straight line, hence it can be assumed that within these limits the percentage of convulsions or symptoms is directly proportional to the logarithm of the dose.

Since the same number of mice were used in each group it is unnecessary to express the numbers of mice convulsing as percentages; the numbers may be used directly. First a measure of the effect of increasing the dose in the ratio 2/1·2 is obtained. The total of mice which convulsed on both high doses is 17+19 = 36, and the mean effect of the high doses is 18. Similarly the mean effect of the low doses is 6. Hence 18—6 = 12 is the effect of multiplying the low dose by 2/1·2.

Next a measure of the difference in the effect of the standard and unknown is obtained. The difference between the effect of the high doses of standard and unknown is 2, while the difference between the effect of the low doses is 4. Hence the mean difference is 3. Since the effects are proportional to log dose, then if M be the log of the ratio of potency of unknown to standard

$$\frac{12}{3} = \frac{\log \frac{2}{1\cdot 2}}{M},$$

or

$$M = \frac{\log 1\cdot 666}{4}$$

$$= 0\cdot 0554.$$

But 0·0554 is log of 1·14, hence the unknown solution was 1·14 times as strong as the solution of standard. A more precise calculation can be made by converting the figures for mice convulsing into probits.

Thus the number of mice convulsing on the high dose of standard was 17 out of 24, or 71 per cent. By turning to Fig. 7, it is seen that the normal deviation corresponding to 71 per cent. is 0·55, hence the probit is 5·55. Likewise the number of mice convulsing after receiving the low dose of standard was 4 out of 24 or 16·6 per cent. The normal deviation corresponding to 16·6 per cent. is seen in Fig. 7 to be —0·97, hence the probit is 5·0—0·97 = 4·03. Similarly for the other proportions.

Or else the probits may be taken direct from tables of probits.

	Response	*Probit*
Standard high dose .	17/24	5·5485
,, low ,, .	4/24	4·0326
Unknown high ,, .	19/24	5·8122
,, low ,, .	8/24	4·5693

Mean probit of high doses is 5·6803
,, ,, low ,, 4·3009

Hence increasing the dose by log 28—log 16·8, which is 0·2218, caused an increase in probit of 1·3794.

The difference of the probits of the high doses of standard and unknown is 0·2637.

The difference of the low dose probits is 0·5367.

Hence the mean difference is 0·4002.

Since 1·3794 is the probit difference due to a dose increase of 0·2218, then 0·4002 is the probit difference due to a dose increase of $\frac{0 \cdot 2218 \times 0 \cdot 4002}{1 \cdot 3794}$ which is 0·0643.

Since this is the log of 1·15

then ratio of $\frac{\text{unknown}}{\text{standard}} = 1 \cdot 15.$

The slope of the line, b, relating log dose to probit increase is given by

$$\frac{\text{mean high dose probit} - \text{mean low dose probit}}{\text{log dose ratio}}$$

or

$$b = \frac{5 \cdot 6803 - 4 \cdot 3009}{0 \cdot 2218}$$

$$= \frac{1 \cdot 3794}{0 \cdot 2218}$$

$$= 6 \cdot 21,$$

and the weight of the result W is given by $W = \bar{B} \times b^2 \times n$, where $\bar{B}$ is the mean of weight factors of individual responses.

Turning to Fig. 7 it is seen that the weight factors are as follows:

17/24 or 71 per cent.,	weight factor	0·57
4/24 or 16·6 per cent.,	,,	0·446
19/24 or 79 per cent.,	,,	0·5
8/24 or 33·3 per cent.,	,,	0·599.

Hence $B = 0{\cdot}529$.

Since n is the number of animals receiving the same dose, we see that

$$W = 0{\cdot}529 \times (6{\cdot}21)^2 \times 24$$
$$= 490.$$

Hence the method enables values for slope and weight to be calculated for each test.

C. The Testing of Preparations of Protamine Insulin Suspensions (with Zinc)

Since the introduction of protamine insulin by Hagedorn, preparations of insulin intended to exert a prolonged action are sold in the form of protamine insulin, protamine zinc insulin, and globin insulin.

The tests which the manufacturer should apply in the course of preparing the solution containing zinc are (1) the ordinary potency test for insulin, (2) a test to ensure that the final product exerts a delayed action, and (3) a test to ensure full combination.

The potency test. The insulin used for preparing the suspension must be tested for potency in the usual way. The strength of the insulin must be such that, when it is diluted with the required amount of the solution which brings about the suspension, the strength of the mixture agrees with the strength stated on the label.

The test for delayed action. Two groups of at least 10 rabbits are injected with test and soluble insulin standard respectively. Blood samples, pooled for each group, are analysed for sugar initially and at each hour for 6 hours. After a suitable interval, preferably not less than 3 days, the groups are crossed over and the test repeated. The final curves, one for test and one for

standard, are compared. When the standard curve returns to its initial level, the test curve must not be above 80 per cent. of its initial level or if the standard curve does not return to its original level in 6 hours, when it reaches 90 per cent. the test curve must not be above 72 per cent. of its initial level. No test of this kind is practicable for globin insulin.

Test for full combination. After adding protamine and buffer to the soluble insulin the supernatant fluid is tested to see that it contains less than 3·85 per cent. of the total potency. This can be done either by a cross-test on 6 rabbits, or a mouse test using 60 mice. It may be noted that a solution of pure crystalline insulin with an activity of 1 unit per c.c. contains about 0·007 mg. nitrogen per c.c. If the supernatant fluid contains as little nitrogen as this, it cannot contain more than the maximum permissible quantity of insulin.

The second and third of the tests are applied to the final product and can be carried out by anyone; the first test, applied to the insulin before it is suspended, can only be carried out by the manufacturer. An approximate estimate of the potency of the insulin present can, however, be obtained by diluting the final product with acid saline and comparing it with the standard insulin in the ordinary way. When diluted (e.g. 20 times) the suspension is broken down.

REFERENCES

HAGEDORN and JENSEN, 1923, *Biochem. Zeit.* **135**, 46; **137**, 92.

HEMMINGSEN, 1933, *Quart. J. Pharm. Pharmacol.* **6**, 39, 187.

HEMMINGSEN and MARKS, 1932, ibid. **5**, 245.

—— 1933, ibid. **6**, 81.

MARKS, 1926, *League of Nations Report on Insulin Standardization.* C.H. 398.

—— 1932, *Quart. J. Pharm. Pharmacol.* **5**, 255.

Quarterly Bulletin of the Health Organization, League of Nations, Special Number, November 1936.

ROWLINSON and LESFORD, 1948, *Quart. J. Pharm. Pharmacol.* **21**, 259.

SOMOGYI, 1930, *J. Biol. Chem.* **86**, 655.

TREVAN and BOOCK, 1926, *League of Nations Report on Insulin Standardization*, C.H. 398.

CHAPTER VI

THE HORMONES OF THE SUPRARENAL GLAND

(*a*) Extract of Suprarenal Cortex

Standard preparation. At the present time there is no standard available for the estimation of extracts of the suprarenal cortex. Attempts at standardization must therefore be confined to comparing one extract with another.

The rat method. While a method described by Bülbring (1937) in which drakes are used is a reasonably accurate method, it is difficult to carry out and involves obtaining large numbers of drakes. It has been superseded by a method devised by Selye and Schenker (1938) of which the details given here are taken from the paper by Vogt (1943).

Groups made up of nine to ten rats of the same age and weight are used. The survival time of rats after adrenalectomy increases with weight and decreases, within certain limits, with the age at which a given weight is attained. Litter mates are evenly distributed among the groups, and the mean weight of each group is made equal. The same ratio of males to females is used throughout the groups.

In one test the rats must be from the same colony. Vigorous strains are more susceptible to cortical hormone than weak strains. In Vogt's work, for one colony of Wistar rats, ages from 23 to 30 days and weights between 38 and 60 gm. proved workable, whereas for another strain (Wistar rats, Dr. K. H. Coward's colony reared under strictly constant dietary conditions) suitable figures were 19–23 days and 38–45 gm. The mean survival time for rats of this last colony was about double that of the first.

Suprarenalectomy is carried out by the lumbar route from a midline incision through the skin. Pentobarbitone can be combined with ether for use as anaesthetic. The rats are given a few drops of 25 per cent. glucose solution soon after the operation. They should eat overnight, usually gain in weight, and be lively and fit the next morning when the experiment is about to start.

Adrenalectomy in the rat. The rat is anaesthetized with ether and the hair along the middle of its back is clipped and moistened with alcohol. An incision is then made along the middle of the back 3–5 cm. long. A small opening is cut in the muscle of the body-wall so that the opening bisects the angle between the lowest rib and the spinal muscles. This opening is about 0·5 cm. By pressing the viscera from below, the kidney is pushed through the opening. The suprarenal gland is then usually visible, connected to the upper pole of the kidney. The ligament on the headward side of the suprarenal, connecting it to the body-wall, is clamped in a small curved artery-clamp and torn away from the body-wall; as it comes away, the suprarenal gland comes with it, and also a piece of the kidney capsule to which it is attached. This completes the removal of the gland; the kidney is pushed back into the abdomen and the opening in the muscle is sewn with a single catgut suture. The gland of the opposite side is then removed through an incision in the muscle of the opposite side. The skin incision is finally sewn with silk.

Performance of test. For the test, the rats are distributed into individual numbered cages early in the morning, given their first injection, and simultaneously placed in a large refrigerator (2°–5° C.). At intervals of an hour and a half further injections are made, their total number being either four or five. The doses are injected as rapidly as possible, all the cages being removed from and replaced into the refrigerator simultaneously. From the fifth or sixth hour after the beginning of exposure, a rapid inspection of the rats is carried out every half-hour, any deaths recorded, and these observations continue till all or nearly all rats have died.

The mean survival time is calculated for each group. When observations have not been continued till all rats in a group have died, the percentage of deaths observed at the half-hourly intervals can be plotted against time on probability paper and the mean survival time found graphically.

Since the rats in each group are not distributed at random, but carefully selected, so as to have in every group representatives of the same litter, ages, weights, and sexes, the test for 'significance' of the result can be based on the standard deviation of the *differences* in survival times between corresponding

pairs of rats in any two groups and not on the standard deviation of the survival times themselves. By dividing the standard deviation of the differences by the root of the number of pairs, the standard error ϵ of the mean difference D can be obtained, and the ratio D/ϵ gives the usual measure for significance, the probability that a result is due to chance being determined by means of Fisher's table (1941).

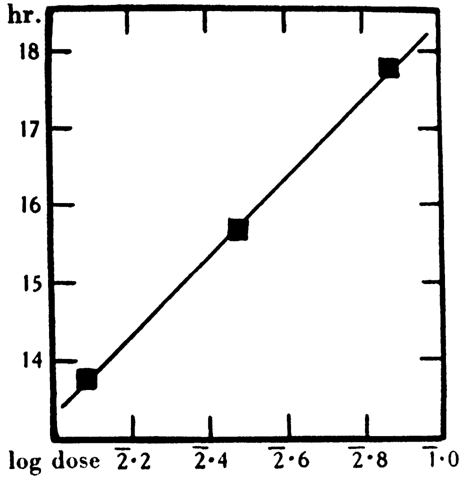


FIG. 29. Dose-response curve. Ordinate: mean survival time (hours) of groups of ten rats. Abscissa: log dose (ml. cortical extract per rat and injection). (M. Vogt, *J. Physiol.* 1943, **102**, 343.)

Fig. 29 is a dose-response curve obtained by injecting 3 groups of 10 rats with increasing doses of a cortical extract ('Eucortone', Allen and Hanburys). Each rat had 5 injections containing 0·012 ml. Eucortone in group 1, 0·03 ml. in group 2, and 0·075 ml. in group 3. In the graph, the logarithms of these doses are plotted as abscissae against the mean survival times (13·75, 15·67, and 17·81 hr.) of the 3 groups as ordinates. As shown in the figure, the 3 points lie in a straight line, and indicate that the response is proportional to the logarithm of the dose. A calculation of the standard errors of the differences in survival times as outlined above led to the following result: the probability that the mean difference in survival time between groups is due to chance is 1:20 for groups 1 and 2, and 1:50 for

groups 2 and 3. In other words, the differences obtained with the three doses, though significant, are not sufficiently large to warrant further reduction of the differences between the doses. The doses had been so selected as to increase by the factor 2·5, and the result of the calculation of the error shows that smaller differences between doses would not have been detected with certainty. This result was confirmed in several other experiments by the observation that the increase in survival time obtained by doubling the basal dose was not significant, whereas that resulting from trebling it was. In this respect the method compares unfavourably with assays on drakes (Bülbring, 1937), in which doses differing by 50 per cent. are distinguishable. Its advantage is its high sensitivity; the total dose required per rat is about 1/25th of that necessary per drake. In Fig. 29, for instance, the smallest dose used for one rat is $5 \times 0{\cdot}012$ or 0·06 ml. extract; the survival time with this dose was 5·75 hours longer than with saline and the response, therefore, considerably above threshold.

If two groups of rats of the same age and average weight are taken from the same colony and treated with the same dose of cortical extract or injected with the same volume of saline in two consecutive experiments with, as far as one can tell, identical environmental conditions, the mean survival time will nevertheless be different. Consequently, in each assay of cortical activity the survival time without treatment must be established for a group of rats injected with saline only, and the susceptibility to hormone by at least one group given a cortical extract used as 'standard'. If quantitative assay of differences between several unknown samples is desired, two groups of rats given two different doses of 'standard' are required.

(*b*) Adrenaline

Adrenaline is submitted to biological test in two circumstances: first, when a sample of adrenaline is to be tested to see if it has the full activity of the pure substance; second, when small amounts of adrenaline which cannot be estimated chemically are to be detected and measured approximately. For the first purpose the test on the blood-pressure of the spinal cat is the best; for the second, the test on the isolated intestine of the rabbit should be used.

Preparation of spinal cat (Dale's method). Full-grown male cats which have not been castrated are the best, but all cats are suitable. The cat is anaesthetized with ether, secured on the operating table, and the trachea is exposed. A tracheal cannula is inserted through which the cat breathes ether from a Woulff's bottle. Clips are put on both carotid arteries temporarily. The cat is turned over on to its belly, the legs are secured, but the head is left free. The next procedure is to expose the spinal cord beneath the long spine of the second cervical vertebra. The skin is divided by a scalpel down the midline from the top of the head to the shoulders. The two flaps of skin are held back on each side by a hook with weight attached. The operator then holds the cat's head in one hand so as to extend the muscles at the back of the neck, and divides the first layer of muscles down the midline from the skull for about 2 or 3 inches. These muscles are held apart by a pair of hooks and weights. The spine of the second cervical vertebra can be felt in the midline. The muscles are cut away from the sides of the spine by cuts made as close as possible to the spine, but these cuts should not be made near the skull end of the spine. The muscle is cleared from the bone by using a thin, blunt dissector, and it is held away by the hooks and weights. The muscle attached to the lower end of the spine is cut through, and the whole of the spine may be cut off with a pair of forceps. At this stage of the preparation the cat should be as deeply anaesthetized as possible in order to minimize the loss of blood. Bleeding from the bone is easily arrested by smearing plasticine over the surface. If bleeding occurs from other tissues, it will always stop if a cotton-wool swab, dipped in hot saline and wrung out, be pressed on the bleeding-point for a minute or two. With a small pair of bone forceps the layer of bone covering the spinal cord is now patiently nibbled away. This must be continued until the dura mater is exposed over a length of about 1 cm., and over its full breadth. The head of the cat is supported by an assistant, and the dura and spinal cord are cut through with scissors. The anaesthetic is removed and artificial respiration is begun. The brain is destroyed by a probe, 3–4 mm. in diameter, thrust through the foramen magnum; the probe is withdrawn and a tapering cone of plasticine is thrust into the brain in its place. Finally, the foramen

magnum is plugged with a small cork. After removing stray portions of plasticine from the spinal canal, the upper end of the cord should appear cleanly cut through, and separated by about 1 cm. from the lower end of the cork. Some bleeding may occur from the end of the cord, but this is soon arrested by applying cotton-wool dipped in hot saline (60° C.). The skin of the back of the neck is sewn together with string, and the cat is turned on to its back once again.

A cannula is put into the left common carotid artery for recording the blood-pressure, and into the right external jugular vein for making injections. When the preparation is made in this way, the initial blood-pressure may be as high as 150 mm. mercury. During the course of 1 hour it usually falls to a level which is often maintained for 24 hours, provided that the temperature of the cat is not allowed to fall. After the completion of the operation, it is best to cover up the cat and to leave it for an hour. The blood-loss during the preparations is variable, and may even alarm the operator, but it is never so great that the cat becomes unsuitable for carrying out the test. With practice it is greatly minimized.

Usually the preparation can be used without delay, but if the blood-pressure is high or variable, the spinal cord should be destroyed by passing a stiff wire down the spinal canal. As the wire is withdrawn the open end of the canal is plugged with plasticine to minimize the loss of blood. The destruction of the spinal cord causes a fall in blood-pressure which may be so great that the cat will die unless it is quickly turned on its back and an injection of adrenaline is made into a vein.

As a standard of comparison, a sample of adrenaline is used which satisfies all tests for chemical purity; the optical rotation should be $-50°$ to $-53°$ in a 4 per cent. solution in N hydrochloric acid. From this adrenaline a solution in weak hydrochloric acid is prepared and diluted with saline to a concentration of 1 in 40,000. This solution must be made acid to litmus with hydrochloric acid, or it will decompose. A similar solution of the unknown is prepared. The first step is to determine the most suitable dose of the standard with which to compare the unknown solution. The dose chosen for the comparison of unknown solution and standard must produce a submaximal effect, yet should not be far from the maximal dose, since in

this region the response to a given dose is more constant than elsewhere. The estimation of adrenaline by this method is very accurate and differences of 6 per cent. can be detected. Fig. 30 shows the comparison of an unknown with the standard in which two injections of the unknown are shown to be greater than a given dose of the standard, and then to be less than a

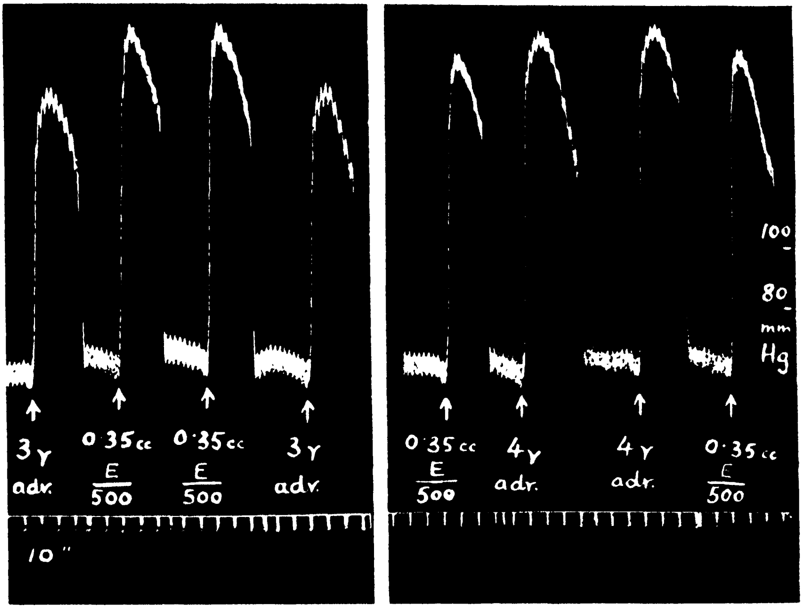


FIG. 30. Comparison of unknown adrenaline solution with a known solution on the blood-pressure of a spinal cat. 0·36 c.c. of the unknown extract (1 in 500 dilution) is greater in effect than 3 μg. adrenaline, and less in effect than 4 μg.

higher dose of the standard. The unknown was an extract *E* of an adrenal tumour. The results can be written:

$$0{\cdot}35 \text{ c.c. } \frac{E}{500} \text{ is greater than } 3\ \mu\text{g. adrenaline,}$$

$$0{\cdot}35 \text{ c.c. } \frac{E}{500} \text{ is less than } 4\ \mu\text{g. adrenaline,}$$

$$\text{therefore} \quad 0{\cdot}35 \text{ c.c. } \frac{E}{500} \text{ is equal to } 3{\cdot}5\ \mu\text{g. adrenaline,}$$

$$\text{or} \quad 1 \text{ c.c. } \frac{E}{500} = 10\ \mu\text{g. adrenaline.}$$

Hence the tumour contained 5 mg. adrenaline (or adrenaline-like substance) per gramme.

Estimation on the rabbit intestine. The estimation by means of the rabbit's intestine is performed by suspending a strip of intestine in an isolated organ-bath. Ringer-Locke (not Tyrode) solution is used.

Ringer-Locke

NaCl	9·0 gm.
KCl	0·42 gm.
$CaCl_2$	0·24 gm.
$NaHCO_3$ . . .	0·5 gm.
Dextrose . . .	1·0 gm.
Distilled water . .	1 litre

The quantity of $CaCl_2$ refers to the anhydrous salt. A stock solution can be prepared from crystalline $CaCl_2$, 6 H_2O, and its strength determined by titration with standard $AgNO_3$. The amount of $NaHCO_3$ is larger than that for ordinary Ringer-Locke which is 0·15 gm. The larger amount gives the solution a reaction corresponding to pH 8·0. The strip of intestine is taken from a freshly killed rabbit and should be 3–5 cm. long. The lever must be weighted with care. The pull on the intestine must not be too great to prevent the appearance of tone in the muscle, and must not be too small to produce a good relaxation when adrenaline is added. The temperature must be kept constant; a temperature about 34° or 35° C. is best. The rate at which oxygen bubbles through the solution must not vary. Fig. 31 shows a comparison of the same adrenal tumour extract with adrenaline as that which was compared on the blood-pressure in Fig. 30. The bath was 100 c.c. capacity. The results were:

$$0{\cdot}2 \text{ c.c. } \frac{E}{500} \text{ was greater than } 1{\cdot}5\ \mu\text{g. adrenaline,}$$

$$0{\cdot}2 \text{ c.c. } \frac{E}{500} \text{ was less than } 2{\cdot}5\ \mu\text{g. adrenaline,}$$

therefore $$0{\cdot}2 \text{ c.c. } \frac{E}{500} \text{ was equal to } 2\ \mu\text{g. adrenaline,}$$

or $$1 \text{ c.c. } \frac{E}{500} = 10\ \mu\text{g. adrenaline,}$$

which is the same result as that on the blood-pressure.

Estimation of adrenaline in a mixture containing *nor*adrenaline (arterenol). The foregoing methods do not distinguish between adrenaline and *nor*adrenaline, though *nor*adrenaline is stated to be more potent than adrenaline on blood-pressure, and less potent as an inhibitor of the rabbit intestine. The relationship varies in different strips of rabbit intestine, as several workers have found. West (1947) has published a table

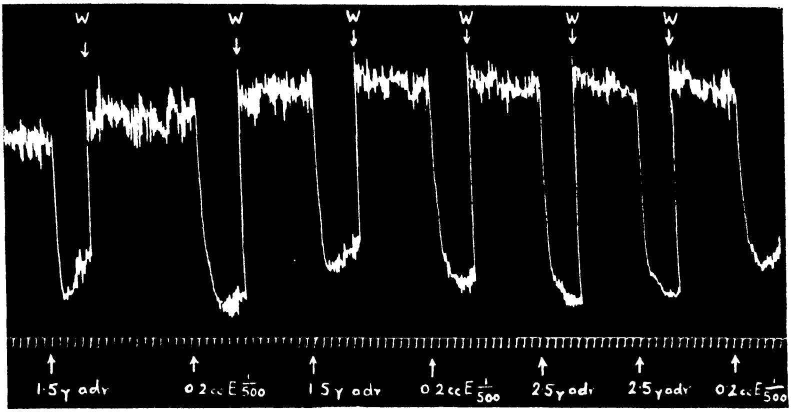


FIG. 31. Comparison of unknown adrenaline solution with a known solution on the isolated duodenum of the rabbit.

showing the order of the ratio in many preparations. This is reproduced in Table XLII. Evidently the best organ to use to measure the amount of adrenaline in the presence of *nor*adrenaline is the non-pregnant uterus of the rat, on which *nor*adrenaline has almost no effect. The frog heart comes next in usefulness.

Rat uterus. The most sensitive biological method of estimating adrenaline is to determine its power to inhibit the action of acetylcholine on the rat uterus. The rhythmic contractions of the uterus are abolished by using Ringer's solution of the following formula proposed by de Jalon, Bayo, and de Jalon (1945).

NaCl . .	9·0 gm.
KCl . .	0·42 gm.
$CaCl_2$ . .	0·06 gm. (¼ of the usual amount)
Glucose .	0·5 gm. (½ of the usual amount)
$NaHCO_3$.	0·5 gm.
Water to .	1 litre

TABLE XLII. *Ratio of dose of dl-noradrenaline to equi-active dose of l-adrenaline*

Test object	*Ratio*	*Excitor (E) or inhibitor (I) action*
Cat, blood-pressure . .	0·8	E
Cat, pregnant uterus . .	0·8	E
Cat, ileum	1·0	I
Cat, nictitating membrane .	1·25	E
Rabbit, ileum . . .	2·0	I
Rat, ileum	3·0	I
Rabbit, pregnant uterus .	4·0	E
Rabbit, non-pregnant uterus .	5·0	E
Frog, blood-vessels . .	5·0	E
Frog, Straub heart . .	8·0	E
Cat, non-pregnant uterus .	10·0	I
Frog, perfused heart . .	33·0	E
Rat, non-pregnant uterus .	100·0	I

A bath of 15 ml. capacity is used. The temperature is maintained at about 31° C., and the bath is aerated with oxygen or oxygen plus 5 per cent. CO_2.

A contraction of the uterus is obtained by adding 5 μg. (or thereabouts) acetylcholine to the bath. The bath is washed out 30 seconds later. Provided the contraction is of suitable height, the same dose of acetylcholine is added to the bath every 90 seconds throughout the experiment and washed out 30 seconds later. The method of estimating adrenaline is to add a known amount of it to the bath 30 seconds before a dose of ACh in order to see how much the effect of ACh is reduced. The dose of adrenaline varies from 0·001 to 0·01 μg. according to the sensitiveness of the uterus. Additions of the standard solution of adrenaline or of the unknown solution can be repeated as soon as the inhibitory effect of the previous dose has disappeared. The amount of the unknown is found which produces the same reduction of the acetylcholine contraction as that produced by a known amount of adrenaline. An example of an estimation is given in Fig. 32. This method was worked out in the Department of Pharmacology, Edinburgh.

Frog heart. A method of estimating adrenaline on the isolated frog heart has been described by West (1943). It has the same accuracy as the blood-pressure of the spinal cat. The

heart is exposed free from pericardium, and the anterior venae cavae are ligatured. A glass cannula filled with frog-Ringer is tied in the posterior vena cava. The heart is removed from the frog and its contractions are recorded on a drum by a thread and a long isotonic lever. The cannula is connected to the

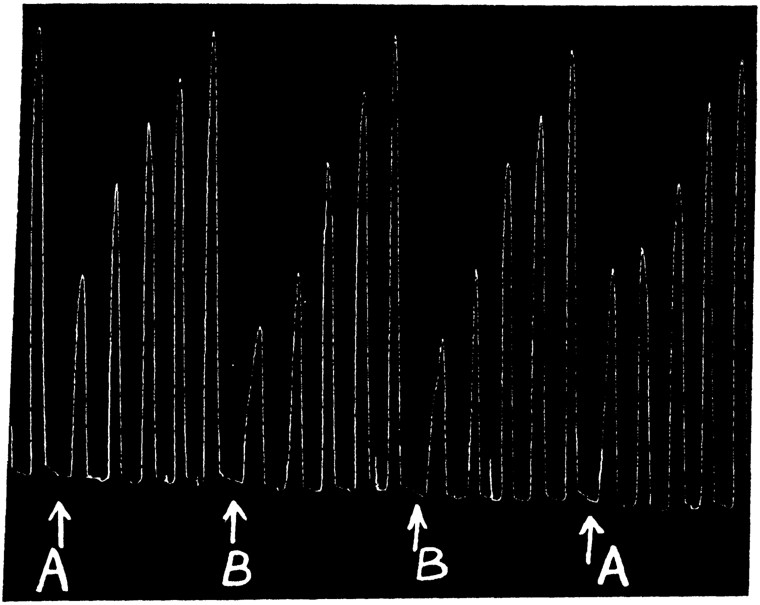


FIG. 32. Estimation of adrenaline on the rat uterus by the inhibition of the effect of acetylcholine. 5 μg. ACh is added to the bath and causes each contraction. At *A*, 0·2 c.c. of an extract was added to the bath 30 sec. before the addition of ACh. At *B*, 0·004 μg. adrenaline was added. The record shows that the effect of 0·004 μg. adrenaline was greater than that of 0·2 c.c. extract. *Nor*adrenaline is almost inactive by this test. (Experiment performed by A. U. Pardoe.)

perfusion apparatus (see Fig. 33) which is supported on a single retort stand. The frog-Ringer solution in reservoir *A* is maintained at a constant level (as shown by an arrow), the adrenaline dilution in bottle *B* being also at this level before perfusion. The two siphon tubes are at the same height above the venous cannula. A comparison will now be described.

Performance of test. A dilution of the standard adrenaline solution was placed in bottle *B*, and tap *B* was opened so that

this solution siphoned over by blowing down outlet tube *B*. Using a slow drum, a short tracing was taken to show the normal beat, and then the three-way capillary tap *C* was turned to allow perfusion of the adrenaline solution in place of the Ringer's solution. This perfusion lasted exactly 1½ or 2 minutes, after which the lever was removed from the drum. The drum

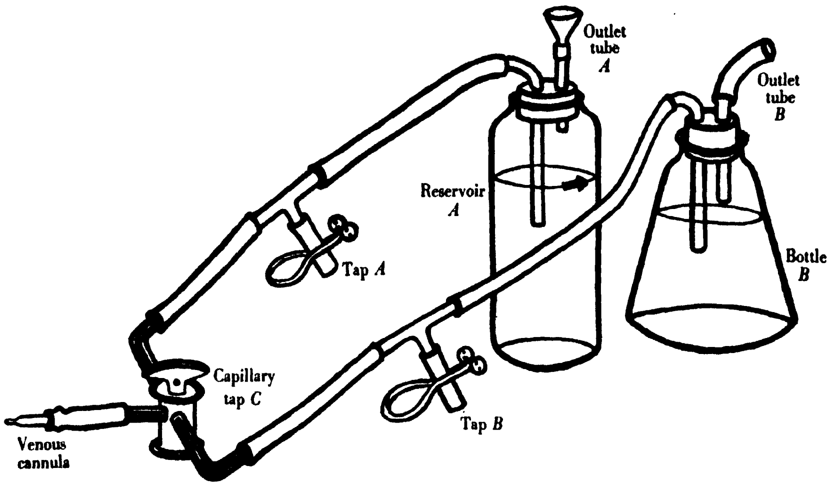


FIG. 33. The perfusion apparatus. The retort stand has been omitted; it supports reservoir *A*, bottle *B*, and the three-way capillary tap *C*. (West, *J. Physiol.* 1943, **102**, 368.)

was stopped, and tap *C* turned to allow the Ringer's solution to reperfuse. The contents of bottle *B* and of the siphon tube *B* were removed, and the liquid remaining in the tube between tap *C* and tap *B* was washed out with Ringer's solution by turning tap *C* and opening tap *B*. The dilution of the test adrenaline solution was then prepared, and this took exactly 4 minutes. During this period the heart-beat had nearly always returned to normal. The drum was started, and the test solution perfused for the same fixed time, at the end of which it was replaced by Ringer's solution. The test adrenalinc was repeated using the same dose, and, finally, the standard so that a group of four results was obtained. It was found that the response to a given dose was more constant in the region of the maximum response than elsewhere, and so solutions were compared by

using doses which just produced submaximal contractions. Equality of response was useful only as confirmatory evidence.

The most accurate procedure for obtaining dilutions was to have the adrenaline solutions supplied in 2 ml. ampoules. Each ampoule was used for one response only, and dilutions were always carried out using one complete set of pipettes and three

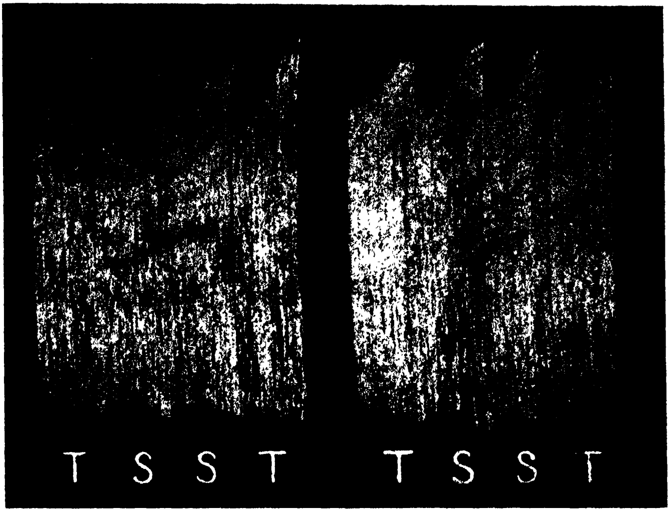


FIG. 34. Effects of standard (*S*) 1/200,000 compared with test (*T*) 1/150,000 and 1/200,000. Each perfusion lasted for 1½ min. only. (West, *J. Physiol.* 1943, **102**, 369.)

100-ml. volumetric flasks. One ml. of the solution in the ampoule was diluted with 99 ml. of freshly prepared boiled and cooled distilled water in flask 1; of this dilution, 1 ml. was removed and diluted with 99 ml. of freshly boiled and cooled distilled water in flask 2; of this dilution, aliquot portions were removed and made up to 100 ml. with Ringer's solution in flask 3. This was the perfusion fluid, and it was used immediately and never more than once, as over 30 per cent. of the activity is lost in 5 minutes. On most hearts, the standard gave an adequate response in a concentration of 1/100,000,000 of adrenaline.

Fig. 34 shows a comparison between this standard (*S*) and a test solution (*T*). From these tracings, (*a*) 1/150,000 of test was stronger than 1/200,000 of standard, or test was stronger

than 150,000/200,000 = 0·75 of standard; (*b*) 1/200,000 of test was weaker than 1/200,000 of standard, or test was weaker than standard. Hence, the approximate strength of the test solution was 87·5 per cent. of the standard, that is, 1/1,143 of adrenaline. The comparison was then repeated using different doses of the test solution so as to obtain closer limits.

The standard error of the test, calculated by the method of Gaddum (1938) using the true values of the 'unknown' dilutions, was 2·17 per cent. The limits of error ($P = 0 \cdot 99$), therefore, are $100 \pm 5 \cdot 59$ per cent. It is recognized that differences of 5–6 per cent. can be accurately measured by the cat method, so that the frog-heart method is just as accurate.

Estimation on the rat colon. Gaddum, Peart, and Vogt have observed (1949) that the action of *nor*adrenaline on the colon of the rat is often greater than that of adrenaline. In some preparations the same effect as is produced by adrenaline is produced by one-tenth the amount of l-*nor*adrenaline. In other preparations the two substances are equally effective. The test on the rat's colon is carried out in the same way and in the same solution as the test on the rat uterus at a bath temperature of 25°–27° C.

Mixtures of adrenaline and *nor*adrenaline. The simplest method of estimating adrenaline and *nor*adrenaline present in an extract is first to use the rat uterus method to determine the amount of adrenaline present (Gaddum, Peart, and Vogt, 1949). Since *nor*adrenaline usually has only 1 per cent. of the effect of adrenaline in this preparation, the *nor*adrenaline can be disregarded.

When the amount of adrenaline in a given volume of the extract has thus been determined, the extract is then injected intravenously into a spinal cat. If adrenaline is injected equal to the amount in the extract, the rise of blood-pressure caused by the adrenaline will be less than that caused by the extract. *Nor*adrenaline is now added to an amount of adrenaline equal to that present in the extract until the rise of pressure caused by the mixture is equal to the rise of pressure caused by the extract. The amount of *nor*adrenaline added is then equal to the amount in the extract.

A better estimate is obtained in the following way. In performing the rat uterus test, a comparison is made of the

relative potency of adrenaline and *nor*adrenaline, and as before the extract is estimated in terms of adrenaline. Two results are thus obtained on the rat uterus; it is found that

$$p\ \mu g.\ \text{adrenaline} = 1\ \mu g.\ \textit{nor}\text{adrenaline}$$

$$1\ \text{c.c. extract} = m\ \mu g.\ \text{adrenaline.}$$

When the rat's colon test is carried out, it is found that

$$q\ \mu g.\ \text{adrenaline} = 1\ \mu g.\ \textit{nor}\text{adrenaline}$$

$$1\ \text{c.c. extract} = n\ \mu g.\ \text{adrenaline.}$$

If 1 c.c. extract contains A μg. adrenaline

1 c.c. extract contains N μg. *nor*adrenaline,

then on the rat uterus

$$A+pN = m \qquad (1)$$

and on the rat's colon

$$A+qN = n. \qquad (2)$$

Subtracting (2) from (1)

$$N(p-q) = m-n$$

$$\text{or } N = \frac{m-n}{p-q}$$

and substituting for N in (1)

$$A+\frac{p(m-n)}{p-q} = m$$

$$\text{or } A = \frac{pn-qm}{p-q}.$$ (See Gaddum and Lembeck, 1949.)

A third method, described by Bülbring and Burn (1949), is to use the nictitating membranes of the spinal cat, one membrane having been denervated 8 days previously by removal of the superior cervical ganglion. *Nor*adrenaline has little action on the normally innervated membrane, but has a similar effect to adrenaline on the denervated membrane. When equipressor doses of adrenaline and *nor*adrenaline are injected, the ratio of the contractions in the two membranes is, for example, 2 for adrenaline, but is 10 for *nor*adrenaline. The effect in the denervated membrane is similar for adrenaline and *nor*adrenaline, but the contraction of the normal membrane is one-half that of the denervated membrane for adrenaline, and one-tenth that of the denervated membrane for *nor*adrenaline.

The ratios for equipressor doses of adrenaline and of *nor*-adrenaline having been determined, the ratios are also determined for mixtures containing

(*a*) 75 per cent. adrenaline and 25 per cent. *nor*adrenaline;

(*b*) 50 per cent. adrenaline and 50 per cent. *nor*adrenaline;

(*c*) 25 per cent. adrenaline and 75 per cent. *nor*adrenaline.

When this has been done it is possible to plot a curve relating the ratio to the composition of the mixture. If the ratio given by the extract is also determined, then from the curve the amounts of adrenaline and *nor*adrenaline in the extract can be found.

This method has been modified by Burn, Hutcheon, and Parker, who determine the ratio of the contraction of the normal nictitating membrane to the rise of blood-pressure. The ratio is in linear relation to the proportion of adrenaline present. The ratio is determined for a series of known mixtures as (*a*), (*b*), and (*c*) above, and then the ratio for the unknown mixture is also determined. The amounts injected should give similar rises in blood-pressure.

REFERENCES

BÜLBRING, 1937, *J. Physiol.* **89**, 64.

BÜLBRING and BURN, 1949, *Brit. J. Pharmacol.* **4**, 202.

FISHER, 1941, *Statistical Methods for Research Workers*, p. 167. London: Oliver & Boyd.

GADDUM, 1938, *Quart. J. Pharm.* **11**, 697.

GADDUM and LEMBECK, 1949, *Brit. J. Pharmacol.*, **4**, 401.

GADDUM, PEART, and VOGT, 1949, *J. Physiol.* **108**, 467.

DE JALON, BAYO, and DE JALON, 1945, *Farmacoterap. Actual*, **2**, 313.

SELYE and SCHENKER, 1938, *Proc. Soc. Exper. Biol.* **39**, 518.

VOGT, 1943, *J. Physiol.* **102**, 341.

WEST, 1943, ibid. **102**, 367.

—— 1947, ibid. **106**, 418.

CHAPTER VII

THYROID

THE estimation of thyroid activity is no longer confined to dried thyroid preparations, for in recent years artificially iodinated proteins, mostly iodinated casein, have been introduced. These iodinated proteins contain a large amount of acid-insoluble iodine, but only a small and varying proportion of it is in combination as thyroxine. Chemical assay has little value, and a fresh importance has been given to biological methods.

One of the oldest methods is the acetonitrile test of Reid Hunt (1905); it depends on the observation that the administration of thyroid to mice increases their resistance to the toxic action of acetonitrile. Gaddum and Hetherington (1931) and Wokes (1935) found it impossible to apply this test with accuracy. Methods depending on observing the oxygen uptake (Gaddum, 1930) or the carbon dioxide output (Mørch, 1929) take a great deal of time, and Kreitmair's (1928) test in which the loss of weight in guinea-pigs is recorded is far from satisfactory.

Survival time in a closed vessel. Recently a test has been described by Smith, Emmens, and Parkes (1947) in which the survival time of mice placed in a closed vessel is observed. The mice die because they use up the oxygen in the vessel, and produce carbon dioxide. If two groups of mice, one of which has been given dried thyroid beforehand, are compared, it is found that the treated mice die much earlier than the control mice. The treated mice have a decreased resistance to anoxia or to carbon dioxide poisoning. Table XLIII shows the difference in the survival time, and the composition of the enclosed

TABLE XLIII (Smith, Emmens, and Parkes)

	Mean survival time, min.	CO_2 *percentage (mean)*	O_2 *percentage (mean)*
Thyroid-treated .	54	10·11	8·56
Controls . . .	175	14·27	4·56

air at the death of the treated mice and the controls. The control mice were treated with casein. The striking feature of the results in Table XLIII is the early death of the thyroid-treated mice at an oxygen percentage 4 per cent. higher and a carbon dioxide percentage 3 per cent. lower than was reached by the controls. It appears that the reduced survival time of the thyroid-treated mice was due in the main to increased sensitiveness to anoxia or to carbon dioxide excess, rather than to an increased metabolic rate.

Performance of the test. Four groups of twenty male mice are used, the weights of the mice being between 18 and 20 gm. Two of the groups receive the substance used as a standard preparation, one dose for one group and a higher dose for the second group. The remaining two groups receive the unknown preparation, one dose for one group and a higher dose for the second, the ratio of the two doses of the standard preparation being the same as the ratio of the two doses of the unknown.

The preparations are most conveniently given by subcutaneous injection of 1 c.c. of solution or suspension. Thyroid preparations which are insoluble are ground in a mortar with 0·9 per cent. NaCl and diluted to give suspensions containing approximately 5 mg. per c.c., or whatever amount is required. Iodinated proteins are prepared for injection by damping 0·5 gm. of the dry substance with sufficient distilled water to make a smooth paste, and 30 c.c. N/10 NaOH is added slowly with continuous stirring for 5 minutes. This suspension is incubated at 37° C. for about 30 minutes or until dissolved. About 50 c.c. distilled water is added and the solution brought to pH 7·4 by adding N/10 HCl, using brom-thymol blue. The volume is brought to 125 c.c. with distilled water so that 1 c.c. contains the activity of 4 mg. Dilutions of this solution are then prepared so that solutions containing 2 mg., 1 mg., and 0·5 mg. per c.c. are obtained. The injections are given on three occasions on alternate days, and the test is performed on the second or third day after the last injection. For example, injections may be given on Monday, Wednesday, and Friday of one week and the test performed on Monday of the next week.

Food is left with the mice until the beginning of the test. Each mouse is then placed by itself in a glass jar of 860–80 c.c. capacity. Convenient vessels in England are known as Kilner

jars which are used for bottling fruit. The glass lid of this jar is fitted with a rubber washer and is secured by a metal screw-band which makes the jar air-tight. The 2-lb. size is suitable. Sawdust is placed in the jars before use.

The experiment is carried out in a room in which the temperature is maintained between 22·5° and 23·5° C., and the

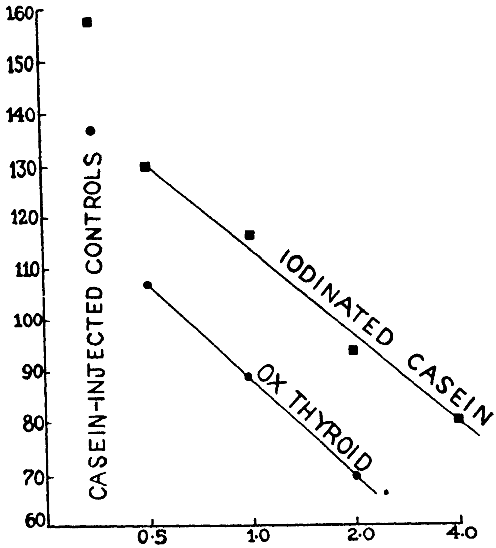


FIG. 35. Relation between survival time (min.) and dose of ox thyroid or iodinated casein (mg.). Each point shows the average for 10 mice. (A. U. Smith, C. W. Emmens, and A. S. Parkes, 1947, *J. Endocrinol.* 5, 186.)

time is noted when the lid is screwed on each jar. The jars are laid horizontally on shelves. After being enclosed the mice are usually very active for 10–20 minutes. Then they become quiet and appear to sleep. There is a progressive increase in the rate and depth of respiration. The terminal symptoms are restlessness for 5–10 minutes, coma, and brief convulsions. The time of the last breath is taken as end point.

Relation of survival time to log dose. Results of experiments in which ox thyroid and iodinated casein were used are shown in Fig. 35, in which the mean survival times of groups of mice receiving a given dose are plotted against the log dose. Of the iodinated casein, solutions were prepared containing 4, 2, 1,

and 0·5 mg. per c.c. and each group of ten mice received three injections of 1 c.c. of one of the solutions on alternate days. A group of control mice received 1 c.c. of a casein solution containing 2 mg. per c.c. A similar experiment was carried out with ox thyroid.

Calculation of result. The calculation of the result from the observations made with two doses of standard and two doses of unknown depends on the linear relation between log dose and survival time. It is made by the method described on p. 19, and illustrated on p. 184.

Smith, Emmens, and Parkes (1947) say that 'with the slope and variation in response encountered in this laboratory, such a test (using eighty mice) should give determinations of relative potency with fiducial limits of error ($P = 0{\cdot}05$) not far exceeding about 70–140 per cent., as long as the criteria for parallel rectilinearity of the two dose/response lines are satisfied'.

REFERENCES

GADDUM, 1930, *J. Physiol.* **68**, 381.
GADDUM and HETHERINGTON, 1931, *Quart. J. Pharm. Pharmacol.* **4**, 183.
HUNT, 1905, *J. Biol. Chem.* **1**, 33.
KREITMAIR, 1928, *Zeit. f. d. g. exp. Med.* **61**, 202.
MØRCH, 1929, *J. Physiol.* **67**, 221.
SMITH, EMMENS, and PARKES, 1947, *J. Endocrinol.* **5**, 186.
WOKES, 1935, *Quart. J. Pharm. Pharmacol.* **8**, 54.

CHAPTER VIII

PARATHYROID EXTRACT

Standard preparation. THERE is no international standard for parathyroid extract, and at the present time there is no dry stable preparation of the hormone available which could be adopted as a standard.

METHODS OF ESTIMATION

Dogs are not suitable animals for a biological method, since it is rarely convenient to house large numbers. It is essential for a good method that the average effect in a group of animals should be determined, and the group must be large enough to reduce the error of sampling to small dimensions. From this it follows that mice or rats, or similar small creatures, are always preferred to larger ones such as dogs. In the second place, the estimation of changes in blood suffers from the drawback that the concentration of many substances in blood depends not only on the liberation of the substance *into* the blood, but also on its removal *from* the blood. When an injection of parathyroid extract is given, calcium is liberated from the bones, causing a rise in the amount in circulation. The rise in the blood calcium is presumably followed by an increased transference of calcium into the cells of different organs, particularly those of the kidney, where it is excreted in the urine. The rise in serum calcium, observed some hours after the injection of parathyroid extract, is, therefore, unlikely to be a measure of the amount of extract injected, since already some of the calcium discharged from the bones will have left the blood for other destinations. To obtain a measure of the effect produced, it will be better to make observations, not on the blood, but on a body-tissue or fluid where the rise in calcium concentration is irreversible, e.g. in the urine. When an injection of parathyroid extract is made into an animal, and an excess of calcium is excreted in the urine, that excess must stand in some relation to the amount of hormone injected.

The rat method. This method depends on the effect of the extract in increasing the output of urinary calcium in rats.

Diet. The procedure used by Dyer (1933) is to take a group of ten rats weighing from 120 to 160 gm. each, and preferably all of one sex. These are prepared by feeding them on a dry diet containing 1 per cent. of calcium carbonate; the effect of parathyroid extract in causing an increased excretion of urinary calcium was found to be much greater when a high calcium diet was used; moreover, if a high calcium diet was given, the increase in excretion was repeated when a second test was performed on the same animals after a few days. The importance of a high calcium diet was suggested by Bülbring (1931), who showed that the effect of parathyroid extract in removing calcium from the bones of rats was much greater if the diet was rich in calcium. The diet used by Dyer is:

Yellow corn (ground)	65 parts
Whole wheat	20 ,,
Dried milk	20 ,,
Casein	9 ,,
Dried yeast	5 ,,
Calcium carbonate	1 part
Sodium chloride	0·5 ,,

When the rats have been fed for a few days on this diet, each rat is put in a metabolism cage, so that its urine can be collected for the estimation of calcium. None of the food must fall through the bottom of the cage into the vessel which collects the urine. If the diet is put in a container outside the cage the rat will scatter some of it inside the cage; if the container is inside the cage the rat will sit on the container, and some urine may be lost. To meet the difficulty Dyer prepares a paste diet consisting of

Butter	10 parts
Compound powder of tragacanth	10 ,,
Diet described above	80 ,,

A large mortar is warmed, the tragacanth put in it, and the butter after being melted is poured in. The butter and tragacanth are mixed and the basal diet added a little at a time. Finally, enough distilled water is added to produce an adhesive mass which does not crumble. A portion of this diet weighing about 10 gm. is made into a ball and placed in the cage twice daily; it is eaten eagerly.

Metabolism cages. Cages for the collection of urine are

readily purchased. As a rule they have a floor of wide mesh through which the urine and faeces fall on to a funnel. At the outlet of the funnel is an inverted pear-shaped glass bulb, which is suspended over a glass collecting vessel. The glass bulb prevents the faeces from entering the collecting vessel, but conducts the urine to it.

The estimation of urinary calcium. In collecting the urine the funnel of each cage is sprayed with distilled water so that the washings run into the collecting vessel. Washings and urine from a group of five cages (containing five rats) are collected together and diluted to 100 c.c. with distilled water. Fifty c.c. is evaporated to dryness on a sand-bath in a 125 c.c. hard glass conical flask, using glass beads to prevent bumping. One to two c.c. concentrated nitric acid is added to the charred residue, and a small funnel is placed in the mouth of the flask, which is then reheated to dryness. The addition of nitric acid is repeated until a white ash remains; sometimes four or five additions are needed, and the process takes 1–2 hours. The white residue is dissolved in 50 per cent. hydrochloric acid, taking from 5 to 10 c.c., according to the amount of residue. To dissolve the residue, heating is necessary, and the heat drives off most of the excess of HCl; the solution is then made up to 10 c.c. For estimation, 5 c.c. are placed in a centrifuge tube (15 c.c. capacity) with a conical end, neutralized to methyl red with 2 per cent. solution of ammonia, 1 c.c. of 4 per cent. ammonium oxalate is added, and the tube allowed to stand overnight. On the next day the tube is centrifuged (10 minutes at 2,000 r.p.m.), so that the precipitate is well packed together at the bottom of the tube. The supernatant fluid is poured out and the tube is inverted on a filter-paper for 5 minutes, when the sides should drain without disturbing the precipitate. The precipitate is washed with 2 per cent. ammonia and collected by centrifuging. The supernatant fluid is again poured away and the tube drained. The precipitate is dissolved in 2 c.c. N sulphuric acid and the tube warmed to 70° C. in a water-bath. Potassium permanganate, 0·01 N, is added drop by drop from a 10-c.c. burette until a permanent pink colour is produced. During the titration the temperature must be kept at 70° C. The 0·01 N solution of potassium permanganate should be prepared fresh when required by diluting a stronger solution. (A normal solution

contains 3·16 gm. per litre; 1 c.c. 0·01 N is equivalent to 0·2 mg. calcium.)

Effect of parathyroid extract on urinary calcium. The effect of injections of parathyroid extract on the output of urinary calcium is illustrated in Table XLIV. The total urinary calcium from 5 rats was determined daily for 4 days; the mean figure was 1·3 mg. daily. Injections of parathyroid extract were then made once a day for 3 successive days, and the urinary calcium during these 3 days and the next day rose to a mean figure of 7·8 mg. daily. During 3 more days the mean figure was 3·8 mg. daily, and when the injections were repeated the mean figure rose once more to 10·4 mg. daily.

TABLE XLIV. *Daily output of* Ca *in the urine from five male rats of total body-weight 730 gm.*

Day	*Injection of parathyroid extract*	*Total Ca mg.*	*Mean daily output*
1	Nil	1·5	1·3
2	,,	0·8	
3	,,	1·9	
4	,,	0·9	
5	0·4 c.c.	6·5	7·8
6	,,	8·5	
7	,,	8·5	
8	Nil	7·7	
9	,,	2·2	3·8
10	,,	4·0	
11	,,	5·1	
12	0·4 c.c.	11·1	10·4
13	,,	17·0	
14	,,	7·2	
15	Nil	6·3	

Method of performing the test. The performance of the test is based on the evidence of Truszkowski, Blauth-Opienska, and Iwanowska (1939) that the relation between the log dose and the excretion of calcium is linear. Four groups of not less than 5 rats of the same sex and similar weight are taken and are given the high calcium diet for a few days before they are put

in the metabolism cages and fed on the paste diet. The total daily calcium output of each group is next determined during 4 days. The rats are then injected during 3 days. When 2 extracts are to be compared, 2 doses of each extract are chosen, the ratio of the 2 doses of one extract being the same as the ratio of 2 doses of the other. Each dose is injected into 1 group of 5 rats. Suitable doses of Parathormone (Eli Lilly) are 0·2 c.c. and 0·4 c.c. for rats of about 150 gm. The urinary calcium is determined for the 3 days of injections and the next day.

The rise in the calcium excretion caused by the injection of the extract is then calculated in each group, and the calculation of the relative potency of the two extracts can be determined from the four doses and the four effects of these doses by the method illustrated for pituitary (posterior lobe) extract on p. 184.

REFERENCES

BÜLBRING, 1931, *Arch. Exp. Path. Pharmak.* **162**, 209.

DYER, 1933, *Quart. J. Pharm. Pharmacol.* **6**, 426.

TRUSZKOWSKI, BLAUTH-OPIENSKA, and IWANOWSKA, 1939 *Biochem. J.* **33**, 1005.

CHAPTER IX

THE OVARIAN HORMONES

I. The Oestrus-producing Hormone

The standard preparations. There are two standard preparations of the oestrus-producing hormone, oestrone and oestradiol monobenzoate. Oestrone, the hydroxyketonic form of the hormone, normally obtained from the urine of pregnancy, was adopted as an international standard in 1932 by the Permanent Commission on Biological Standardization. The material deposited in the National Institute for Medical Research, London, was supplied mainly by Girard, but also by Laqueur, Doisy, and Butenandt. The different samples were mixed together and the mixture was distributed in ampoules which were kept for 10 weeks *in vacuo* over phosphorus pentoxide. The ampoules were filled with dry nitrogen and sealed.

In the years following the adoption of this standard, esterified forms of the hormone were prepared and came into regular use. When these were tested in comparison with the standard oestrone, the relative potency was found to vary according to the solvent and according to the species employed in the test. In 1935 the decision was therefore taken at a second conference to establish a second standard which should be the monobenzoate of the dihydroxy form of the hormone, to be known as oestradiol monobenzoate. The material was provided by Laqueur, Girard, and Butenandt; this was mixed and distributed in ampoules with the same precautions as were taken for the first standard.

The unit. When oestrone was adopted as a standard the unit was defined as the specific oestrus-producing activity in 0·0001 mg. of the standard. When oestradiol monobenzoate was adopted as a second standard the unit for the benzoylated preparations of the hormones was also defined as the activity present in 0·0001 mg. of the standard, but this was not intended to indicate that the activity present in 1 unit of the two standards is the same.

Preparation of solutions for injection. Solutions of the oestrus-producing hormones are prepared in water and also in oil; when

a test is to be performed the standard preparation must be dissolved in the same solvent as that in which the unknown preparation is dissolved. The standard preparation can be kept conveniently in alcoholic solution in which it is stable, as shown by the work of Rowlands and Callow (1935). A 0·1 per cent. solution in 97·5 per cent., or in absolute alcohol, is prepared most easily by warming the alcohol before pouring it on the crystals which have been weighed in a watch-glass. The crystals must be rubbed for a long time to dissolve them. A watery solution is then prepared by adding the required volume of alcoholic solution to a much larger measured volume of water. To make an oily solution the oil is weighed into a beaker; it cannot be measured by volume, but the volume is calculated from the specific gravity; 9·15 gm. olive oil is equivalent to 10 c.c. The alcoholic solution is added to the oil, and the alcohol is driven off by warming on a water-bath. When a solution in oil is injected into mice, the volume of injection should be 0·1 c.c.; when a solution in oil is injected into rats, the volume of injection should be 0·2 c.c.

Estimation of activity. There are two methods of determining the relative potency of preparations of the oestrus-producing hormone, namely, by the vaginal smear method and by the method in which the uterus is weighed. In both methods ovariectomized animals are required. The animals are anaesthetized with ether. An incision through the skin 0·5–1·0 cm. long is made. As the ease of the operation depends on making this incision in the right place the spot must be selected with care; it should be midway between the last rib and the knee, and about 1 cm. lateral to the spinal muscles. Having incised the skin, a parallel but shorter incision is made through the muscles. Usually a mass of fat is seen immediately underneath, inside the peritoneal cavity. This is withdrawn through the hole by a pair of forceps, and in it is found the ovary, which may be recognized by its attachment to the uterus. The ovary is usually red or reddish-yellow, though sometimes it is colourless. The ovary is drawn through a loop in a silk ligature, and this is tied so as to enclose the whole ovary and as much of the adjacent fat as possible. The tissue tied off by the ligature is cut away, and the end of the uterus is returned to the peritoneal cavity. One or at most two catgut sutures are sufficient to

unite the cut edges of the muscle, and two or at most three silk stitches are sufficient to unite the skin. The ovary of the other side is then removed through a separate incision. The complete operation of removing both ovaries need not take more than 6–7 minutes. It is wise to put the animals in a warm place when the operation is complete; then they rapidly recover.

If the vaginal smear method is to be used, it is wise to remove not only the whole ovary but also the adjacent portion of the uterus; there is then less fear of ovarian tissue regenerating. For this test at least forty animals are required.

A. The Smear Method

The original method of taking a smear was to rub a glass rod or a metal spatula with a smooth round end along the vaginal wall, and then to transfer the cells so collected to a drop of water on a microscope slide. A better method is to put a drop of water in the vagina, and then to take it out for examination. Small tubes are prepared fitted at one end with a rubber teat and drawn out at the other end to a hole 2 mm. diameter. The edge of the hole must be smooth. A drop of tap-water is taken up in a pipette of this kind and transferred to the vagina of the rat or mouse; it is then taken out again and examined under the low power of the microscope on a glass slide. There are three types of cell to be seen:

(*a*) The polymorpho-nuclear leucocyte.
(*b*) A much larger, round nucleated cell.
(*c*) A large squamous epithelial cell often called a cornified cell.

In the smear of the ovariectomized mouse or rat, when no injection of the hormone has been given, there are usually large numbers of leucocytes; there are in addition a few nucleated cells and occasionally a few squamous cells.

The changes which characterize oestrus. The epithelial surface of the vagina of the ovariectomized animal is thin and consists of very few layers. In consequence leucocytes readily penetrate the epithelium and are found on the surface. Under the influence of the hormone the epithelium becomes much thicker; the nucleated cells multiply vigorously, and those nearest the surface lose their nuclei and form the large squamous

cells. Through this barrier of nucleated cells and squamous cells leucocytes cannot penetrate.

These changes take approximately 48 hours to reach their full development from the time of the first injection of the hormone. The period during which the thick squamous epithelium is maintained depends on the animal and on the amount of the hormone, but if this is not unusually large, the period lasts a few hours only. Unless the vagina is examined during this period, the full development of oestrus is not observed although it may have occurred. Since the examinations of the vagina cannot be made repeatedly in routine work, the observer must decide not only what animals are actually in oestrus, but also what animals are likely to pass into oestrus soon, and what animals have recently passed through oestrus.

Before oestrus has occurred a drop of water placed in the vagina will be found to contain either nucleated cells only or a mixture of nucleated and squamous cells. During oestrus the drop will contain squamous cells only. After oestrus the drop will contain squamous cells and many leucocytes which appear as the squamous epithelium separates from the vaginal wall.

Performance of test. When injections are made in oily solution the dose is given in one injection. When injections are made in watery solution the dose is often given in three injections at 4-hourly intervals, or in four injections at 12-hourly intervals. When the hormone is injected in watery solution it is believed to be excreted rapidly, and for this reason several injections are given; the actual interval between the injections and their number must be decided by each worker for himself; provided always that the arrangement of injections for the unknown preparation is the same as that for the standard preparation. Injections are made under the skin and not into the muscles.

If the injections of oily solutions are given early on day 1, then the positive reactions occur mainly on day 3, though a few may be seen on day 4. Examination of the vaginal contents must be made twice on day 3 and early on day 4. If the injections are given at 12-hourly intervals in aqueous solution, the reactions occur on the fourth and fifth days. The most definite positive response which can be observed is the appearance of squamous cells only in the drop from the vagina. The most usual positive response is indicated by a mixture of

nucleated and squamous cells in the first examination of day 3, and a mixture of squamous cells and leucocytes in the second examination. A worker may also decide to count a response positive if he observes nucleated cells only in both examinations on day 3, or if he observes a mixture of nucleated and squamous cells in both of these examinations. The feature in common of all positive responses is the disappearance of leucocytes; yet this disappearance is relative rather than absolute, for search will usually discover a certain proportion of leucocytes among the cells in the drop of water.

When different workers speak of a positive response, it follows that they do not always mean the same thing; a response which is positive in the eyes of one is negative in the eyes of another. Some refuse to consider a response positive unless the whole length of the vaginal canal is covered with squamous epithelium, since it is said that copulation does not occur in the normal animal otherwise. Some are content to regard a response as positive provided that rapid inspection of the drop under the microscope shows a preponderance of cells not present in the vagina of the uninjected animal, even though a small proportion of leucocytes is present. These differences account for the wide variations in the amounts of different compounds which are equivalent to '1 mouse unit' or '1 rat unit'.

Arrangement of animals for a test. For a test 40 animals are required, 20 of which receive a dose of the standard preparation, and 20 receive a dose of the unknown. The groups should be arranged so that for every animal of a given weight in one group there is an animal of a similar weight in the other group; it is then unnecessary to adjust the dose according to the body-weight. The animals in the one group each receive the same dose of the standard preparation, and those in the other group receive the same dose of the unknown, and the proportion of animals in each group in which a positive response occurs is observed. Sometimes preliminary observations are made in order to determine a suitable dose of the unknown preparation to inject into twenty animals; very often, however, the results on a few animals are misleading, and little time is saved by these observations. Some workers do not use animals for a test more than once in 14 days. If they are used at shorter intervals, the response to an injection is often increased by the effect of the

previous injection. Provided that the standard preparation is injected into half the animals at every test, this effect of the previous injection can be neglected and animals may then be used once a week.

Dosage of oestrone. The amount of oestrone which contains activity equal to one rat unit or one mouse unit varies greatly according to different workers. The variations are so great that it is clear that the expression 'rat unit' or 'mouse unit' is entirely without meaning. Different figures are given in Table XLV. These are only a few which have been published and are selected by chance.

TABLE XLV

		Amount of oestrone (μg.) equal to 1 rat or mouse unit	
	Animal	*Water*	*Oil*
Curtis and Doisy . .	Rat	0·3	. .
Burn and Elphick . .	,,	0·6	1·0
D'Amour and Gustavson .	,,	0·74	1·3
Hain and Robson . .	,,	2·5	3·3
Butenandt and Hildebrandt	Mouse	. .	0·125
Marrian	,,	0·06	. .
Rowlands and Callow .	,,	0·047	. .
Burn and Elphick . .	,,	0·36	. .
Schoeller, Dohrn, Hohlweg.	,,	. .	0·5

The dose-effect curve. The relation between the dose of the hormone and the percentage of animals in which a positive response occurred was determined by Coward and Burn (1927). Their observations were made in ninety rats and seventy mice, and they found the same relation between dose and percentage response for both species. D'Amour and Gustavson (1930) obtained a closely similar curve, and also Rowlands and Callow (1935). The correspondence between the curve of Coward and Burn and that of Rowlands and Callow is interesting because the former workers injected a very crude preparation in an oily emulsion, while the latter injected crystalline oestrone in watery solution. The curve is shown in Fig. 36 in which the dose which gives a response in 50 per cent. of animals is given the value 1.

The cross-test. In making a comparison between an unknown

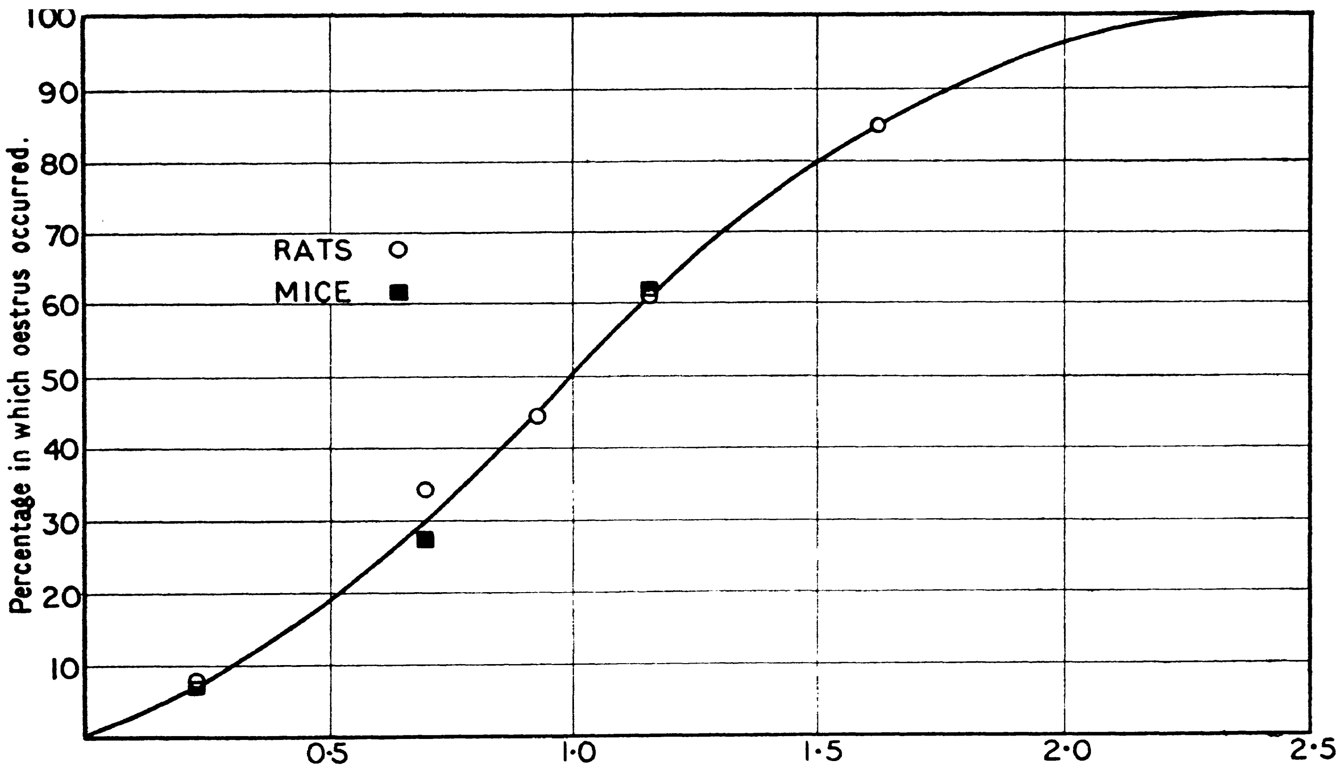


FIG. 36. Relation of dose of oestrus-producing hormone to the percentage of mice or rats in which oestrus occurs. The doses are relative and not actual doses.

preparation and the standard, Laqueur proposed that a test should be performed in two parts; in the first part one group of animals receives the unknown preparation and another receives the standard; in the second part the groups are reversed. An example of the testing of some unknown crystals in comparison with the international standard gave the following results. Solutions of the standard and also of the unknown were prepared in 97 per cent. alcohol, each containing 1 mg. in 25 c.c. From each solution in alcohol a solution in olive oil was prepared containing 0·001 mg. in 0·2 c.c. The solution of the standard was injected into twenty rats and the solution of the unknown was also injected into twenty rats, the dose for each rat being 0·2 c.c. The results of a cross-test carried out on 2 days with an interval of 1 week between are given in Table XLVI.

TABLE XLVI

Day	*Material injected*	*Group of rats*	*Proportion in oestrus*
1	Standard	A	2 out of 20
	Unknown	B	6 out of 20
2	Standard	B	12 out of 20
	Unknown	A	10 out of 20

Hence the injection of the standard caused oestrus in 14 out of 40, or 35 per cent., and that of the unknown in 16 out of 40, or 40 per cent. From Fig. 36 the doses giving 40 per cent. and 35 per cent. responses are in the ratio of 0·86:0·77. Hence the potency of the unknown preparation is 0·86/0·77 of that of the standard, or 1·12 times that of the standard. Hence the unknown preparation contains 11,200 units per mg.

B. The Method in which the Uteri are Weighed

Bülbring and Burn (1935) have described a method of estimation which depends on the increase in the weight of the uterus which the hormone causes in ovariectomized rats. They take young female rats weighing about 40 gm., choosing litters in which there are three or four according to the number of groups which will be used. The ovaries are removed under ether anaesthesia, care being taken not to excise the adjacent portion

of the uterus. Two days later those animals which are to be injected receive the first dose subcutaneously in 0·2 c.c. olive oil; the injection is repeated on the 3 following days and then 2 days after the last injection the animals are killed and the uteri dissected. Thus, if the day of operation is the first day, injections are given on the third, fourth, fifth, and sixth days, and the animals are killed on the eighth day. The dissected uteri are fixed in a modified Bouin's solution consisting of saturated picric acid 750 c.c., formaldehyde 150 c.c., glacial acetic acid 100 c.c., and urea 10 gm. After 24 hours the uteri are dried between pieces of filter-paper and weighed. To ensure uniformity the dissection of the uteri is done by one person. The weight of each uterus was calculated per 100 gm. body-weight of rat. Table XLVII shows the relation of dose of crystalline oestrone to the increase in the weight of the uterus. These and other results show that within the range of doses given there is a simple linear relation between the logarithm of the dose and the mean increase in weight.

TABLE XLVII.

Figures are weights of uteri in mg. per 100 gm.

	Dose injected daily for 4 days			
Litter	*Nil*	*0·2γ*	*0·4γ*	*0·8γ*
1	42	..	124	..
	..	..	100	..
2	28	106	116	145
3	26	72	88	135
4	31	42	68	115
5	..	64	111	136
6	28	70	111	133
7	..	56	68	85
8	36	42	63	87
9	38	65	70	150
	..	63	72	..
Average	33	64	90	123

Methods for detecting small amounts of oestrogens. Hartman, Littrell, and Tom (1946) have described a method of testing oestrogens in young guinea-pigs weighing about 200 gm. The membrane closing the vagina disappears as oestrus approaches;

it is regenerated after 4 days and remains intact for the remaining 12 days of the cycle. Ovariectomized adult guinea-pigs can be used as well as the young ones.

The dose is administered in two injections, one half on each side of the membrane closing the vagina. The needle is inserted at the tip of each horn of the crescent-shaped depression and is directed towards the midline.

When the membrane is intact the vagina is dry and pale. A few hours after the injection of oil or blood containing an oestrogen the area appears congested. Frequent observations should be made, which reveal a pinpoint opening at any place of the crescent; this increases in size and becomes a slit, or remains a pinpoint. The slit is seen for 3–18 hours before the vagina opens completely. If complete opening occurs there is mucus secretion and vaginal cornification.

The time of opening is related to the amount of oestrogen injected. Using oestradiol dipropionate and giving injections in 0·02 c.c. peanut oil, the mean time of opening was 10 hours when the dose was 20 μg., and was 96 hours when the dose was 4×10^{-4} μg.; this threshold dose achieved a pinpoint opening and not a complete opening.

Chick oviduct method. Dorfman and Dorfman (1948) have described a chick oviduct method which is not very sensitive, but which has a high degree of accuracy. The experiments were mostly made on 4-day-old White Leghorn chicks, though some were 7-day-old. The material studied was injected subcutaneously once daily for 5 days and the weights of the oviduct and of the chick were determined 24 hours after the last injection. The total dose was dissolved in 0·5 c.c. corn oil and 0·1 c.c. was given at each injection. The chicks were killed with chloroform, the oviduct dissected, blotted on a towel, and weighed to 0·5 mg. on a torsion balance. The results were expressed as the ratio of the weight of the oviduct in mg. $\times$ 100 to the body-weight in gm. Oestrone, oestradiol, oestradiol benzoate, stilboestrol, and other oestrogens were active. The mean 'oviduct ratio' for groups of 15–25 chicks injected with corn oil as controls varied from $9\pm0{\cdot}86$ to $11\pm0{\cdot}87$. Stilboestrol in a dose of 25 μg. increased the 'oviduct ratio' to $29\pm1{\cdot}8$ in a group of 28 chicks; in a dose of 1·6 mg. stilboestrol increased the 'oviduct ratio' to $564\pm26{\cdot}6$. The authors say: 'At

concentration levels of 25 μg. and 50 μg. of stilboestrol using a simplified statistical design, the potency ratio can be determined with standard errors of ± 10 per cent. and ± 7 per cent. respectively when totals of 20 and 40 chicks were employed. Similar high accuracy was obtained for certain concentrations of oestrone and oestradiol benzoate.'

II. Estimation of Material containing Progesterone

Standard. At an international conference held in London in 1935 it was recommended that progesterone be adopted as the international standard. This standard is distributed from the National Institute for Medical Research, London.

The unit. The recommendation was made at the international conference that the unit be defined as the specific activity present in 1 mg. progesterone.

Method of comparison. A simple method of comparison is that in which immature rabbits are used. The method was outlined by Clauberg (1930, 1931), but the quantitative application was elaborated by McPhail (1934). Progesterone has, by itself, little or no effect on the uterus of the immature rabbit. If extracts of corpora lutea have an effect, this is due to the presence in them of oestrogenic substances. When injections of oestrogens are given, the uterus develops from the infantile form (Fig. 37, 1) to that seen in oestrus (Fig. 37, 2); the diameter increases from three to four times, but the endometrium remains compact; there are a few glands evenly distributed throughout the stroma, and the epithelial lining of the lumen is almost unbroken by glands. If progesterone is injected when the uterus has developed in this way, the typical progestational proliferation occurs. The degree of development depends on the dose, and the different stages are shown in Fig. 37, 3–6; they will be described later. McPhail has described a scale in which different numbers correspond to different degrees of progestational proliferation; his method then consists in determining the scale number which represents the mean degree of proliferation produced in a group of rabbits by a given total dose of material. At the time his paper was published no substance suitable as a standard was available. Now that progesterone has been isolated, all estimations of unknown extracts must be made in comparison with progesterone.

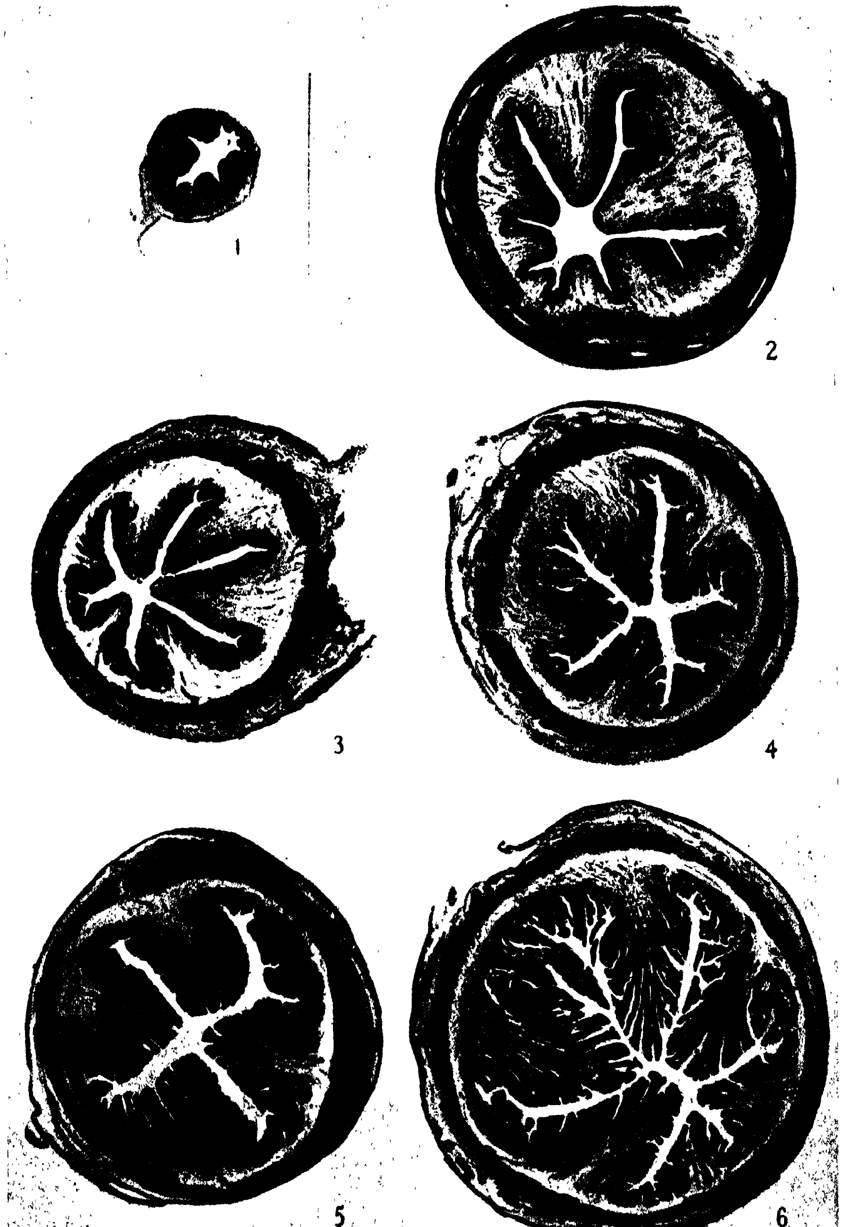


Fig. 37. Sections of uterus: 1 = normal infantile uterus; 2 = uterus after injection with oestrin; 3–6, uteri taken from rabbits injected with oestrin and varying amounts of progesterone. (McPhail, *J. Physiol.* 1934, 83, 145.)

Animals used. Rabbits weighing 750–950 gm. should be chosen for preference; McPhail's extreme weights were 700–1,120 gm. It is sometimes difficult to distinguish the sex of young rabbits, and care must be taken to see that males are not included in the group.

Injections. The injections of oestrone are given over a period of 6 days; they may be given once daily for 6 days, or they may be given on the first, third, and fifth days. The dose is 25 units daily, or 50 units (= 5 μg. oestrone) on each of the 3 days just mentioned; it is injected in oil in a volume of 0·2 c.c. into the muscles of the hind leg; if the first injection is made into the left leg, the next is made into the right, and so on. Thus the total dose of oestrone is 150 international units.

The injections of the unknown material are given on the seventh, eighth, ninth, tenth, and eleventh days, one injection being given daily. A suitable total dose of progesterone is 0·5 mg. distributed in daily doses of 0·1 or in 0·2 c.c. oil. If it is important to save time the injections can be given on 3 days only, instead of on 5 days, though the total dose should be the same; the results of McPhail are almost the same for the two periods.

Preparation of sections of uteri. The rabbits are killed on the day after the last injection. There are different methods of examining the uteri; they may be cut fresh with a freezing microtome by the Schultz-Braun's method of rapid examination of tumour tissue; this method is described below. An alternative is to fix small pieces of tissue in warm formalin for 10 minutes and then to freeze, but there is a risk of the central part of the section dropping out. The slower methods involve embedding in gelatin or paraffin, but these can be greatly accelerated by the use of dioxane (diethylene dioxide) (Weissberger, Young, and Carleton, 1934) as a substitute for alcohol and clearing reagents before embedding in paraffin. In dioxane, small thin pieces of tissue become water-free in 2–3 hours. For histological methods, the book by Carleton (1925) should be consulted.

Rapid method of preparing sections. An ordinary freezing microtome (Leitz) is used. The circular freezing-table is replaced by one which has a straight edge (as a chord of the circle) which must be parallel to the final position of the knife. There is also a knife-cooler, the jet of which is screwed on to the knife.

The rabbit is killed, the uterus removed, and a piece 5 mm. long cut out. This is placed on the freezing-table close to the straight edge so that the section is made transversely. If the uterus is very thin, several pieces from different parts of the horn can be placed side by side to support one another; they are cut simultaneously. It is advisable to put a piece of wet filter-paper underneath to hold the block of tissue more firmly. The tissue and the knife are now frozen. Care must be taken not to freeze the tissue too hard or the sections will crack. The knife should be covered with frost; if there are drops at the sides, it is too warm. If the tissue is too warm the sections collapse before they can be removed. When the right degree of cold is obtained, the cut is made through the tissue, stopping, however, about 1 mm. from the edge, so that the tissue is not quite cut through. Generally the section rolls up, and must be smoothed out with a fine camel's-hair brush. A clean microscope slide, free from grease, is rested against the jet of the knife-cooler and lowered close to the section spread out on the knife (without touching the knife-edge) until the section attaches itself to the glass. At the same instant the section is cut right through. It is then fixed in osmic acid vapour, that is, the slide is placed in a closed vessel in the bottom of which is some filter-paper wet with osmic acid. Some glass beads are placed on the paper so that the slide does not come in contact with the acid. The section sticks to the glass by its own protein. The preparation is placed for a short time in 60 per cent. alcohol and can then be stained with haematoxylin. It is washed and examined under the microscope.

Degree of proliferation. During the early phases of progestational proliferation glands begin to develop round the lumen, but they are quite short and the deeper stroma is unaffected. This is the degree of proliferation which McPhail calls 1 (see Fig. 37, 3). As development proceeds the glands increase in number and length; the deeper stroma is still unbroken and stains lightly; this is degree 2 (see Fig. 37, 4). Further development results in the glands extending and penetrating deeper into the stroma, which in cross-section is becoming much reduced in amount. The glands are still narrow and the glandular tissue has lost its appearance of being a border round the lumen; this is degree 3 (see Fig. 37, 5). Finally the glands

penetrate nearly to the muscular layer, and the stroma remains as the centre of each endometrial fold. The glands are distended, and the endometrium appears broken up; this is degree 4 (see Fig. 37, 6).

The quantitative comparison. In making a comparison of an unknown extract with the standard progesterone, at least

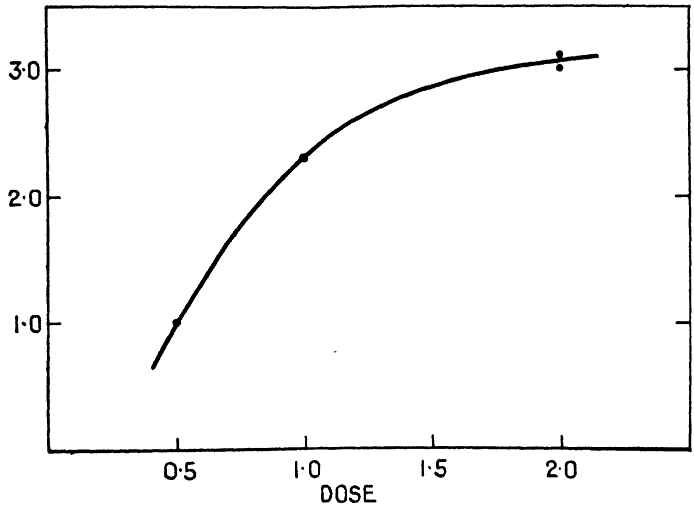


FIG. 38. Relation between dose of progestin (relative scale) and the mean degree of proliferation (ordinate) in the endometrium of the rabbit's uterus. (McPhail, *J. Physiol.* 1934, **83**, 145.)

5 rabbits should be injected with the unknown extract and 5 rabbits with the standard. The 5 rabbits receiving the unknown extract are injected with the same dose; similarly, the rabbits receiving the standard receive the same dose. The uteri of all rabbits are then examined and the mean figure for the degree of proliferation in rabbits injected with the unknown extract, and also the mean figure in rabbits injected with the standard are found.

The relative potency of doses causing different degrees of proliferation is obtained from Fig. 38, which is based on the curve in McPhail's paper. Suppose that a total dose *d* c.c. of an unknown extract produces a mean degree of proliferation = 1·5, while a total dose of 0·5 mg. progesterone produces a mean degree of proliferation = 2·5, then Fig. 38 shows that the

relation of doses producing degrees of proliferation in the proportion of 1·5 to 2·5 is as 0·65 to 1·15. Hence

$$d \text{ c.c. unknown extract} = \frac{0{\cdot}65}{1{\cdot}15} \times 0{\cdot}5 \text{ mg. progesterone}$$

$$\text{or 1 c.c. extract} = \frac{0{\cdot}28}{d} \text{ mg. progesterone}$$

$$= \frac{0{\cdot}28}{d} \text{ units.}$$

Estimation of small amounts of progesterone. For the estimation of tiny amounts of progesterone, Hooker and Forbes (1947) have proposed a method in which the material to be examined is injected into the cavity of the horn of the uterus of an ovariectomized mouse. Mice are used from which the ovaries have been removed at least 16 days before. The injection of progesterone results in nuclear hypertrophy in the endometrial stroma.

The volume injected must be very small, for if as large as 0·01 c.c., the uterus is distended and the tunica propria is obliterated. For the injection a tuberculin syringe of 0·5 c.c., of which the piston is driven home by a micrometer screw, is used. Advancing the screw by 0·001 inch delivered 0·0003 c.c., as determined by filling the syringe with mercury and weighing what was delivered. The volume injected was always 0·0006 c.c.

The mouse is anaesthetized with sodium amytal and one uterine horn is pulled out of a ventral midline incision. A ligature is tied tight round the ovarian end and a loose ligature is passed round the horn 5 mm. lower down. These ligatures are held vertical by artery-clips attached to a horizontal rod placed above the mouse. A needle, passing under the horn, and resting on the mouse, prevents the horn from going back into the abdominal cavity. Care is taken that the blood-supply to the entire uterus is preserved. The horn is grasped with forceps about 3 mm. lower than the loose ligature and the needle of the micrometer syringe is pushed into the lumen just in front of the forceps and continued until near the tight ligature. The injection is made, and as the needle is withdrawn the loose ligature is tightened and tied; in this way leakage of the injected material is avoided.

The mouse is killed 48 hours after injection. The horn is fixed in Lavdowsky's fluid (Williams and Hodge, 1943); after which paraffin sections 6 microns thick are stained with haematoxylin and eosin. A positive result consists in enlargement of the stromal nuclei of the endometrium. The nuclei assume a smooth slightly elongated oval outline. Chromatin particles are fine and evenly distributed and the nucleolus is conspicuous. It is necessary to distinguish between stromal nuclei and those of the glandular epithelium which are not modified by progesterone. Doses as small as 0·0002 μg. progesterone have been found to cause a positive response. In examining a section it is necessary to have sections known to be negative and known to be positive at hand for purposes of comparison. The reaction is not caused by injecting desoxycorticosterone acetate or testosterone.

REFERENCES

Bülbring and Burn, 1935, *J. Physiol.* **85**, 320.
Burn and Elphick, 1932, *Quart. J. Pharm. Pharmacol.* **5**, 192.
Butenandt and Hildebrandt, 1931, *Zeit. Physiol. Chem.* **199**, 243.
Carleton, 1925, *Histological Technique.* London: Oxford University Press.
Clauberg, 1930, *Zbl. Gynak.* **54**, 7, 1154, 2757.
—— 1931, *Die weiblichen Sexualhormone.* Berlin: J. Springer.
Coward and Burn, 1927, *J. Physiol.* **63**, 270.
Curtis and Doisy, 1931, *J. Biol. Chem.* **91**, 647.
D'Amour and Gustavson, 1930, *J. Pharmacol. Exp. Ther.* **40**, 473.
—— 1936, ibid. **57**, 472.
Dorfman and Dorfman, 1948, *Endocrinol.* **42**, 85.
Hain and Robson, 1936, ibid. **57**, 337.
Hartman, Littrell, and Tom, 1946, ibid. **42**, 85.
Hooker and Forbes, 1947, ibid. **41**, 158.
McPhail, 1934, *J. Physiol.* **83**, 145.
Marrian, 1931, *J. Soc. Chem. Ind.* **50**, 368.
Rowlands and Callow, 1935, *Biochem. J.* **29**, 837.
Schoeller, Dohrn, and Hohlweg, 1935, *Klin. Woch.* **23**, 826.
Weissberger, Young, and Carleton, 1934, *Lancet.* **227**, 1279.
Williams and Hodge, 1943, *Anat. Rec.* **87**, 181.

CHAPTER X

MALE HORMONES

The active principles. THERE are many chemical substances which possess male hormone activity, and at least three which occur naturally. The first isolation of a crystalline substance from male urine was carried out by Butenandt (1931), who gave it the name androsterone. Ruzicka and his co-workers (1934) prepared it synthetically from *epi*cholestanol. In the same year Butenandt and Dannenbaum isolated dehydroandrosterone from male urine, which Ruzicka again prepared by degradation of cholesterol. In 1935 David, Dingemanse, Freud, and Laqueur obtained a crystalline substance from testis which they called testosterone, and Butenandt and Hanisch produced it artificially from dehydroandrosterone.

The practical importance for standardization of the existence of these different substances is that they are qualitatively dissimilar in physiological properties. For example, both androsterone and testosterone will cause growth of the capon's comb, but if doses which have the same effect on the comb are injected into castrated male rats the effect of the two substances on the accessory sexual organs, particularly on the seminal vesicles, is very different (Deanesly and Parkes, 1936). Preparations of male hormone from urine generally contain both androsterone and dehydroandrosterone. At the international conference held in London in 1935 it was agreed that all comparisons with androsterone should be made by the capon's comb method. This is accordingly the first method to be described below. The circumstances in which the rat can be used are not yet clear, for the test is not strictly specific; probably comparisons with testosterone can be made in this way; nevertheless the test on the rat is simple and is, therefore, described as well.

The standard preparation. The international standard is androsterone crystalline and artificially prepared with melting-point 183·5°–184·5° C. (corrected) and a rotation $(\alpha)_D^{20°} + 97{\cdot}3$ in absolute alcohol. It was mainly contributed by Miescher of the CIBA works (Basel). The finely ground crystals were

distributed into small glass tubes in which they were dried *in vacuo* over phosphorus pentoxide; the tubes were then filled with nitrogen and sealed. Each tube contains about 50 mg. and is stored in the National Institute for Medical Research, London, at a temperature of $-2°$ C. in the dark.

Solutions are prepared for a test by dissolving a weighed quantity of the crystals in oil; since it has been shown that the oily solution is stable even when kept at 150° C. for 24 hours, it is safe to use heat in dissolving the crystals.

The essential requirement for the use of the standard is that both the standard and the extract which is to be compared with it must be administered in identical solvents; the division and spacing of the doses and the method of giving them must also be the same. The international conference was of opinion that artificially esterified preparations of male hormones cannot be properly compared with the standard preparation of androsterone.

The unit. 'The international unit of male hormone activity' is the 'activity of 0·1 mg. of the international standard preparation of androsterone, as measured by a specific biological reaction'. Koch has found that testosterone contains 1 unit of activity in 13–16 μg.

A. The Capon's Comb Test

The birds which have been most used for this test are Brown Leghorn capons. Different breeds differ in their suitability; thus Callow and Parkes (1935) found that the effect of a given quantity of androsterone injected daily is much less in the Plymouth Rock than in the Brown Leghorn. Brown Leghorns are, however, large birds; they occupy much room in a laboratory and eat correspondingly. Parkes (1936) has pointed out that the Old English Game Bantam is much smaller and therefore much more convenient to house. Fully grown birds often weigh no more than 500 gm. The operation of castration in the adult Leghorn is very difficult, so that it is usually carried out in birds 1–2 months old; it is then necessary to wait about 6 months before the birds can be used for the test. The Bantam on the other hand can be castrated when adult, and can be used 2 months later. In buying adult Bantams it is essential to see that they have entire combs; many have their combs cut off.

Method of operation. The bird is anaesthetized with ether and stretched on its side on a board with the legs pulled backwards and the wings held together upwards. The feathers which cover the last two ribs are plucked, and a skin incision is made between them in front of the head of the tibia. By inserting the blunt points of a pair of scissors the muscle between the last two ribs is divided, and the ribs are held apart by a pair of small retractors. The operator, wearing a head-lamp to illuminate the field, then looks for the testis, which is a white body 1–1·5 cm. long and 0·3–0·5 cm. in diameter in a young bird. The testis is found close to the midline of the posterior abdominal wall, beside the vena cava. To approach it, coils of intestine must be pushed on one side, and these are best kept out of the way by small pads of cotton-wool. The testis is covered with a transparent capsule and the success of the operation depends on opening this capsule. If the attempt is made to remove the capsule as well as the testis, bleeding will occur; if, however, the capsule is opened, using long pointed forceps and fine scissors, the testis is easily removed in one or two pieces without the use of ligatures. If the testis is crushed in removing it, small pieces may remain behind. The opening between the ribs is closed by a ligature around the last two ribs, the skin is sewn up, and the operation is then performed on the other side.

To castrate adult Bantams a ligature is tied around the attachment of the testis. The attachment is smaller than in many breeds, and if care is taken to see that the ligature is really tight, the testis can be removed whole. The stump of the efferent ducts and the pedicle should be thoroughly cauterized. The skin wound only is closed.

Measurement of the comb. Various methods have been used for measuring the growth of the comb. The simplest method is to measure with a pair of dividers and to take the sum of the length and the height. Some workers take photographs of the comb, and measure the size of the photograph with a planimeter. Others have recorded the size electrically by interposing the comb in the path of light falling on a photo-electric cell. Gallagher and Koch have compared the method of direct measurement with measurement of photographs, and have found that direct measurement is more reliable.

The injections. A suitable procedure for injections is to make

injections once a day for 5 days, giving 0·2 c.c. of an oily solution by injection into the pectoral muscles. The effect depends on the volume of oil injected; a given quantity of hormone injected in a volume of 0·2 c.c. has been found to exert a greater mean effect than when injected in 1·0 c.c. The dose is given irrespective of body-weight.

A suitable total dose is 10–20 units, that is to say, 2–4 units a day for 5 days. The number of injections can, of course, be varied, but must be the same for the unknown preparation as for the standard.

The dose-response curve. Greenwood, Blyth, and Callow (1935) carried out an experiment in which they used 5 groups of 5 Leghorn capons each; the different groups received a total dose of 0·5, 1·0, 2·0, 4·0, and 8·0 mg. androsterone respectively. One-fifth of these totals was given daily for 5 days in 0·2 c.c. arachis oil. The relation of the mean increase in each group to the dose is shown in Fig. 39*a* in which the total dose is measured as abscissa and the comb-increase after 5 days (length+height) is measured as ordinate. In Fig. 39*b* the logarithm of the dose is used, and the points then lie on a straight line.

On the basis of these results a comparison of an unknown preparation with the standard can be made with 2 groups of 5 birds each. The average standard deviation of the response of 25 birds was found by Greenwood, Blyth, and Callow to be 2·1 mm., hence if 5 birds are used in a group the standard error is $\pm 1{\cdot}0$ mm., and a difference should be 3·0 mm. before being regarded as significant. A suitable comb-increase ($L+H$) is about 10 mm., which was given in these experiments by a total dose of 2 mg. of androsterone (= 20 units). If the relation between dose and the increase in the size of the comb has not been determined in a laboratory, then 4 groups of 5 birds each should be used to compare an unknown preparation with the standard. Two doses of standard should be chosen, and 2 doses of the unknown, the ratio of the 2 doses of standard being the same as the ratio of the 2 doses of unknown. The relative potency of unknown and standard can then be determined as on p. 184.

Interval between tests. After a series of injections have produced a comb-growth, the comb regresses when the injections stop. If the growth has been more than a few millimetres the comb takes a long time to return to its former size, but it is

unnecessary to wait for this before performing another test. If the period between two tests is not less than 26 days, the first test does not greatly affect the second. Thus Greenwood, Blyth, and Callow found that a total dose of 2 mg. androsterone pro-

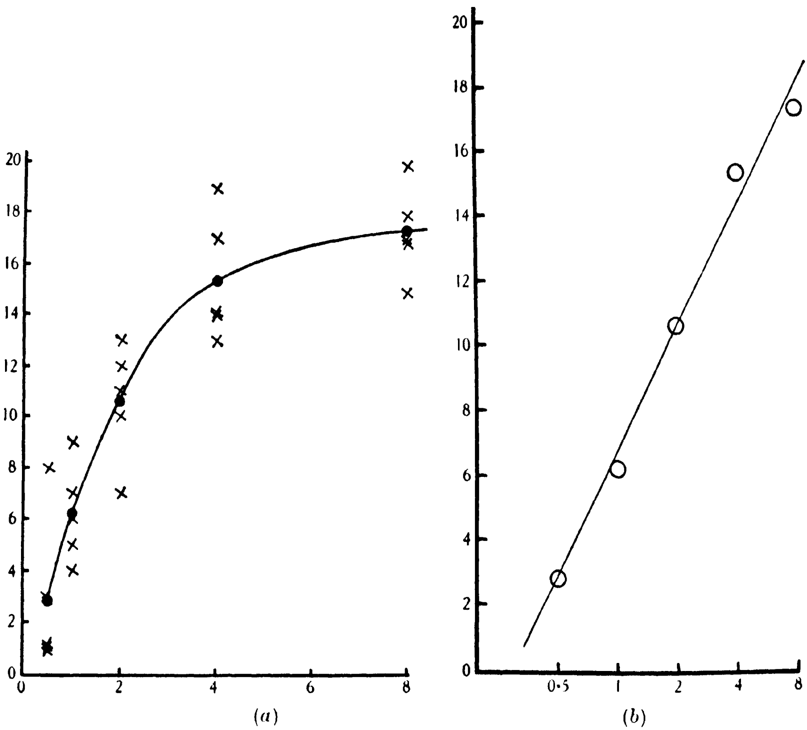


Fig. 39. (*a*) Relation of dose of androsterone (abscissa) to comb-growth (length and height in mm.); (*b*) relation of logarithm of the dose to comb-growth. (Greenwood, Blyth, and Callow, *Biochem. J.* 1935, **29**, 1400.)

duced a mean growth of 10·6 mm. in five birds in one test, and a mean growth of 12·8 mm. in the same birds in a second test begun 26 days after the end of the first. Gallagher and Koch lay stress on the effect of earlier tests, and when the unit is defined as the amount of hormone necessary to produce a certain comb-growth (e.g. 5 mm.), the effect of earlier tests may be important. If, however, the test is a comparison in which the unknown sample is compared with the standard, and in which all birds have been used at the same time in different tests, then

the effect of earlier tests can be disregarded provided an interval of 26 days comes between two tests. This interval is less when Bantams are used.

Age and weight of birds. The comb-response to a given dose of hormone is not affected by the age of the birds; whether they are 4 months or 6 years old the mean response is the same. Doses should not be given according to body-weight, for the mean effect of a given dose in birds weighing 1,500 gm. was not significantly different from that in birds weighing 3,000 gm.

Initial comb-length. Gallagher and Koch consider that in recording comb-growth a correction should be made for the initial length of the comb. They consider length only, and not length+height, and make a correction of 0·2 mm. for every mm. by which the initial comb-length differs from 57 mm. If the initial comb-length was 62 mm., and if the comb-growth was 10 mm. in a test in which five daily injections were given, then the corrected comb-growth would be 9 mm. Greenwood, Blyth, and Callow have failed to observe any relation between response and initial comb-size. Light is stated to affect the comb-response; birds injected with the unknown sample must, therefore, be kept in similar conditions of lighting to those injected with the standard.

B. The Rat Method

The method consists in taking a series of young male rats, of removing the testes, and of determining the effect of the injection of the hormone upon the weight of the prostate and seminal vesicles.

The operation of castration. Young rats weighing 30–40 gm. are anaesthetized with ether and laid on their backs on a board; the hind legs are tied to the sides of the board. The hair over the scrotal area is clipped, and a transverse cut 1 cm. long is made through the skin of the scrotum. When the sides of the incision are pulled apart the testes can be seen. One is gripped firmly in a pair of toothed forceps, and is separated from surrounding tissue by pushing this away with a blunt dissector. The testis lies inside the transparent tunica vaginalis; a cut at the end enables the tunica to retract and the testis to come out of it. A silk ligature is tied around the tunica vaginalis, interrupting the blood-supply to the testis and leaving the whole of

the testis and adjacent fat to be cut off close to the ligature. The second testis is then removed in the same way. The skin incision is sewn up with two or three silk ligatures. During the days following castration the prostate and seminal vesicles diminish steadily in size, and the rats should not be used for an experiment until there is little further change in size. An interval of 10 days is probably sufficient, though many workers prefer to wait for 30 days.

Injections and arrangement of test. In comparing two preparations the best arrangement is to have one group of uninjected animals, two groups injected with doses of the one preparation, and two groups injected with doses of the second preparation. If, however, each group is to contain a rat taken from one litter, it is possible to use those litters only which have five male rats. Hence in making the comparison the worker must often be content with four groups of animals. The doses are usually given in oil, once daily for not less than 5 days; Deanesly and Parkes (1936) make injections for 10 days. These workers have demonstrated the importance of the volume and composition of the oil used as a solvent for androsterone and testosterone. When 10 mg. of androsterone were given as a total dose in 5 c.c. arachis oil, the mean weight of the prostate was 128 mg., whereas when given in 1 c.c. arachis oil, the mean weight was 66 mg. In another experiment they compared the effect of a total dose of 10 mg. androsterone given in a volume of 2 c.c. olive oil with the same dose in the same volume of arachis oil, castor oil, and propylene glycol (a solvent which is miscible with both oil and water). The mean weights of the prostate were 78, 133, 149, and 199 mg. respectively for the four solvents in the order given. These differences are large; they make it clear that the composition of the solvent is of great importance. Finally, Deanesly and Parkes (1936) demonstrated the increase in the activity of testosterone produced by using arachis oil containing 10 per cent. palmitic acid. A total dose of 2 mg. in 5 c.c. arachis oil produced a mean prostate weight of 183 mg.; when the 5 c.c. arachis oil contained 0·5 gm. palmitic acid the weight was 252 mg.

In routine tests a worker must fix the volume and composition of the oil he uses; a suitable daily volume is 0·5 c.c. injected in different places on different days.

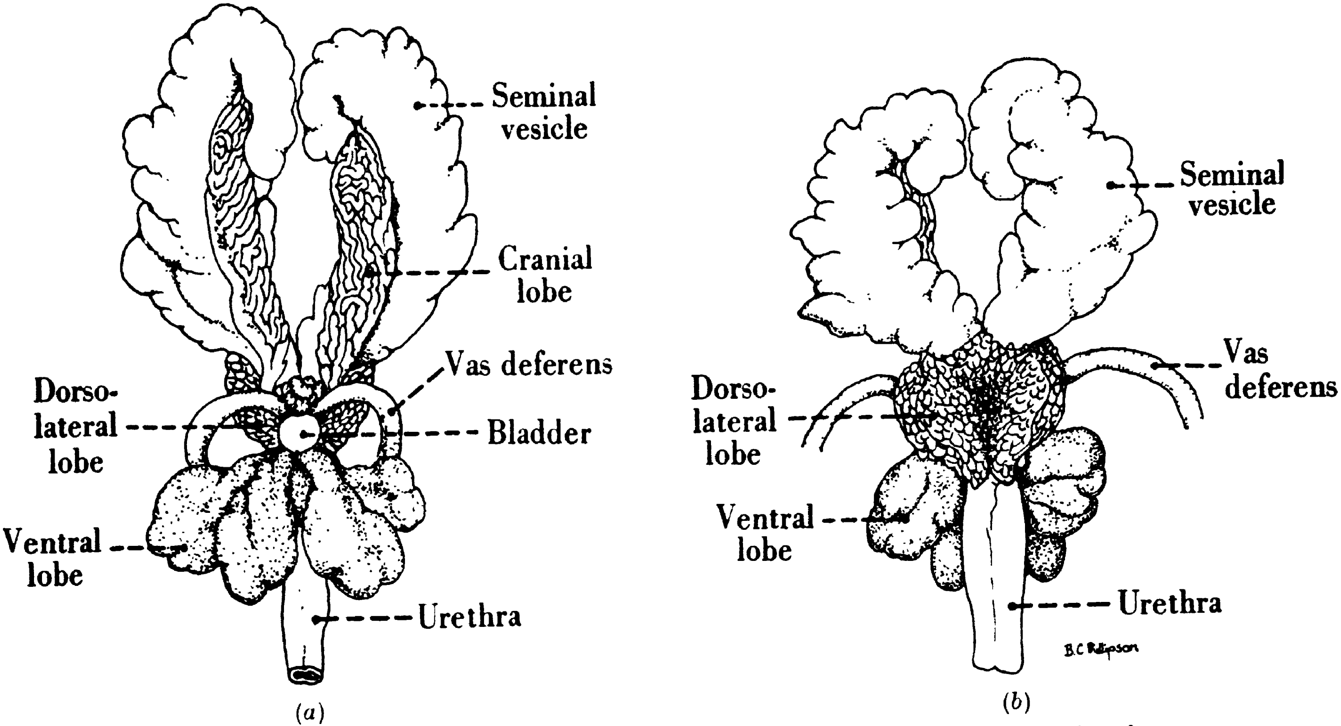


FIG. 40. To show the prostate and seminal vesicles of the rat: (*a*) ventral, (*b*) dorsal. (Callow and Deanesly, *Biochem. J.* 1935, **29**, 1424.)

Dissection of the organs. The following description of the prostate and seminal vesicles of the rat is based on Callow and Deanesly's paper, from which Fig. 40 is reproduced. The prostate consists of (1) cranial lobes or coagulating glands attached to the length of the seminal vesicles; (2) dorso-lateral lobes round the base of the seminal vesicles and almost encircling the urethra; (3) ventral lobes attached only by their ducts. The dissection is best practised in large rats, in which the different parts are easily seen; it can then be attempted in small rats. When a worker is in doubt whether tissue is prostatic or not, he should look at it under the microscope; the convolutions of the prostate are unmistakable; particular care is required not to remove the ventral lobes. To weigh the organs they are dissected free from fat as shown in Fig. 40, and fixed overnight in Bouin's fluid (see p. 248). While portions of fat remain, the tissue floats in the fluid, but when they have been removed it sinks. On the following day the tissue can be pressed between pieces of filter-paper to dry it; then, after removing the pieces of vas deferens and as much of the urethra as possible without interfering with the prostate, the tissue can be weighed. The weight is the combined weight of prostate, seminal vesicles, and the included portion of the urethra. Alternatively the tissue can be transferred from Bouin to 70 per cent. alcohol; the cranial prostate lobes can be dissected from the seminal vesicles, these can be freed from the dorso-lateral lobes of the prostate and removed to be weighed. By dissecting from below, the ventral and dorso-lateral lobes of the prostate can be freed from the urethral musculature; the different portions of the prostate can then be weighed together.

Relation between dose and effect. The relation between the logarithm of the dose and the effect (*a*) on the prostate, (*b*) on the seminal vesicles, is shown in Fig. 41 (*a*) and (*b*). The relation is given for androsterone and also for testosterone. The relationships are not linear but sigmoid. The curves for the effect of the two substances on the prostate are roughly parallel, which suggests that the difference in their action is only quantitative. The curves for the effect of the two substances on the seminal vesicles are quite dissimilar, which shows that the difference in the action is qualitative as well as quantitative.

Estimations of smaller amounts of androgen. In the adult

capon the daily dose of testosterone required to cause growth of the comb is about 14 μg. given intramuscularly. The application of the hormone directly to the comb requires smaller

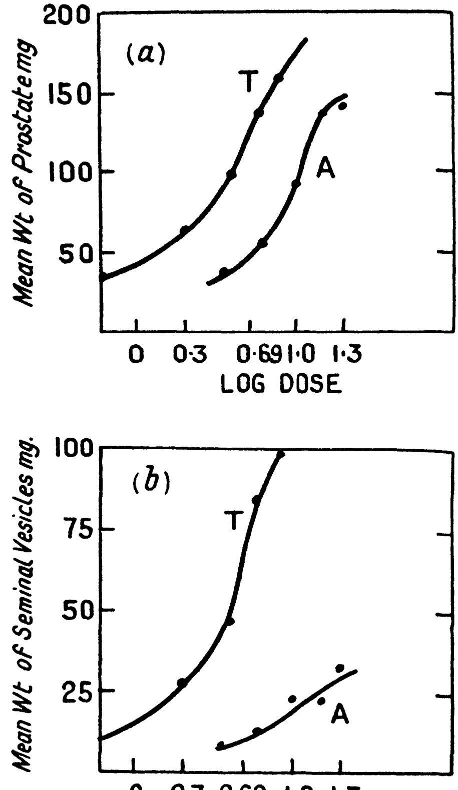


FIG. 41. (*a*) The relation between log dose of testosterone, *T*, and of androsterone, *A*, to the growth of the prostate of the rat; the curves are parallel. (*b*) The relation to the growth of the seminal vesicles; the curves are not parallel. (Deanesly and Parkes, *Biochem. J.* 1936, **30**, 291.)

amounts, and Frank, Klempner, Hollander, and Kriss (1942) have described a method in which White Leghorn chicks, 2–3 days old, are used. The material dissolved in sesame oil is applied once daily for 7 days and the chicks are killed by chloroform 24 hours after the last application. For each application 0·05 c.c. is spread over the entire surface of the comb by a

syringe needle. The material is not rubbed in and the comb is touched as little as possible.

The comb-size is determined after death by excising and weighing the comb. To excise it two parallel longitudinal incisions are made along the base of the comb at the junction with the scalp. The incisions go vertically down to the skull. Then the comb is undercut by one incision starting from behind. After removing blood under the comb by touching it on a towel, the comb is weighed. Comb-weights vary from 16 to 165 mg.

The simplest use of this method will be to choose two doses of the standard in a given ratio to one another, and two doses of the unknown in the same ratio, the doses of the unknown being expected to produce the same effects as the doses of the standard. The effects are expressed as mg. comb-weight per 100 gm. body-weight. The calculation can be made as on p. 184.

Results with this method have been published by Dorfman (1948). Using 16–20 chicks in a group, and working in the range of 20–40 μg. androsterone as the total dose per chick, he found a mean experimental error of 13 per cent.

REFERENCES

BUTENANDT, 1931, *Zeit. angew. Chem.* **44**, 905.

BUTENANDT and DANNENBAUM, 1934, *Zeit. physiol. Chem.* **229**, 192.

BUTENANDT and HANISCH, 1935, *Ber. deutsch. Chem. Ges.* **68**, 1859.

CALLOW and DEANESLY, 1935, *Biochem. J.* **29**, 1424.

CALLOW and PARKES, 1935, ibid. **29**, 1414.

DAVID, DINGEMANSE, FREUD, and LAQUEUR, 1935, *Zeit. physiol. Chem.* **233**, 281.

DEANESLY and PARKES, 1936, *Lancet*, London, **230**, 837.

DORFMAN, 1948, *Endocrinol.* **42**, 1.

FRANK, KLEMPNER, HOLLANDER, and KRISS, 1942, ibid. **31**, 63.

GALLAGHER and KOCH, 1935, *J. Pharmacol. Exp. Ther.* **55**, 97.

GREENWOOD, BLYTH, and CALLOW, 1935, *Biochem. J.* **29**, 1400.

PARKES, 1936, *Quart. J. Pharm. Pharmacol.* **9**, 669.

RUZICKA, GOLDBERG, MEYER, BRUNGGER, and EICHENBERGER, 1934, *Helv. chim. Acta*, **17**, 1395.

RUZICKA and WETTSTEIN, 1935, ibid. **18**, 986.

CHAPTER XI

ANTERIOR LOBE OF THE PITUITARY GLAND

International standards. THE international standards at present available are Thyrotrophin, Prolactin (which is also called Galactin or Mammotrophin), Chorionic gonadotrophin, and serum gonadotrophin. The units are the activity in 0·1 mg. Prolactin, in 0·1 mg. Chorionic gonadotrophin, and in 0·25 mg. serum gonadotrophin. There is no unit defined for Thyrotrophin.

I. THE THYROTROPHIC HORMONE

The two methods in use for estimating this hormone are the histological method and the method in which the increase in weight of the thyroid is observed. To attempt to give a quantitative expression to a degree of histological change is, of course, possible, but laborious and only approximate; it is simpler and more accurate to determine the weight change. Some workers have formed the opinion that the weight change does not run parallel to the change in the histological picture, and therefore do not use the weight method; probably this view is due to the use of too few animals for the histological method. Certainly it is highly improbable that there are two factors in the anterior lobe which act on the thyroid, one of which is responsible for the weight change and the other for the histological change. If there is only one factor, then it is preferable to estimate its potency by the more accurate method.

The method of Rowlands and Parkes (1934). Female guinea-pigs as nearly as possible 200 gm. in weight are used. Rowlands and Parkes found that when guinea-pigs of this weight were killed, their thyroids, after being dissected and fixed in Bouin's fluid overnight, had an average weight of 31 mg., and that as the animals grew the mean increase was 1 mg. for each 6 gm. body-weight. The figure of 31 mg. varies according to the animals, and is probably affected by the method of fixing. Thus in another laboratory five guinea-pigs weighing 225–260 gm. (the mean being 238 gm.) were found to have thyroids weighing 20–27 mg., the mean being 22·6 mg.

The method of estimation is derived from the relation shown in Fig. 42 between the log of the daily dose of anterior lobe extract and the mean increase in weight of the thyroids of groups of guinea-pigs. In the experiment shown in Fig. 42, 10 guinea-pigs were used in each group and doses of 2, 4, 7, and 10 mg. were given once daily for 5 days to the different groups. The injections were made under the skin and the animals were killed on the day after the last injection. The mean weights of the thyroids increased in proportion to the dose, and when related to the logarithm of the dose lay on a straight line.

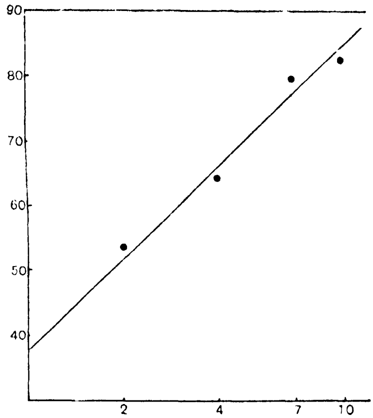


FIG. 42. Relation of dose mg. (logarithmic scale) of anterior lobe extract (abscissa) to mean weight of thyroid of the guinea-pig, mg. (Rowlands and Parkes, *Biochem. J.* 1934, **28**, 1829.)

The comparison of two preparations. The comparison is made by taking four groups of female guinea-pigs of weight about 200 gm. Two of the groups receive daily doses of the standard, the dose for the one group being twice the dose for the other. The unknown preparation is also given to two groups, the dose for one group being twice the dose for the other.

Five animals are used in one group, and injections are given under the skin once daily for 5 days. The thyroids are dissected on the sixth day. The figures in Table XLVIII illustrate the variation which is seen within a group.

TABLE XLVIII

Uninjected animals		*Injected animals*	
Body-weight gm.	*Thyroid-weight mg.*	*Body-weight gm.*	*Thyroid-weight mg.*
235	24	205	73
240	22	220	56
260	27	205	57
225	20	220	41
230	20	200	46
Mean 238	22·6	210	54·6

Histological methods. Those who are interested in histological methods should consult the papers by Aron (1930, 1932), *Loeser (1932)*, Sevringhaus (1933), and Heyl (1933).

II. Prolactin

The pigeon method. Riddle, Bates, and Dykshorn (1933) have described a method of using ring-doves for the estimation

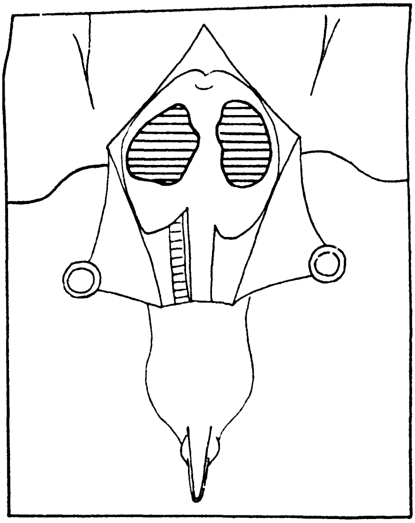


Fig. 43. Showing position of crop glands in the pigeon; the crop has been opened ventrally.

of prolactin by means of its action in stimulating the growth of the glands which occur in the crop of both male and female birds, and which produce a milky secretion. When resting, these glands appear as little more than corrugations in the mucous membrane; but if stimulated by the injection into the pigeon of an active extract they grow into easily defined structures. The diagram in Fig. 43 shows the position of the glands.

In their paper, Riddle and his colleagues give charts which show the relation between the size of the crop glands of single birds and the dose of hormone injected. Unfortunately the weight of the crop glands and the dose are not the actual weights and doses, but are corrected figures from which it is difficult to discover the simple relation between dose and mean

effect, which is needed for comparing two preparations. Dyer (1936) has carried out experiments on groups of pigeons weighing from 220 to 530 gm., the mean weight being about 350 gm. He overcame the difficulty of differences in body-weight by arranging the birds so that the total weight in one group was similar to the total weight in another. Thus in four groups of six birds the group weights were 2,290, 2,145, 1,950, and 2,240 gm. All birds in one group received the same dose without reference to body-weight. The doses were injected three times daily for 4 days into the breast muscles. On the fifth day the birds were killed, the crop glands were dissected and, after being pressed between filter-paper, weighed. The results obtained in one experiment are given in Table XLIX.

TABLE XLIX. *Figures are the weight (gm.) of the glands from one bird.*

Total daily dose	*Nil*	*0·25 c.c.*	*0·5 c.c.*	*1·0 c.c.*
	0·3	1·3	1·8	3·9
	0·4	1·4	2·0	4·2
	0·8	1·6	2·1	4·8
	0·9	2·2	2·9	4·9
	1·0	3·7	3·4	5·2
	1·5	3·8	5·9	12·6
Mean . . .	0·8	2·3	3·0	5·9

The figures in Table XLIX show the great variation in the weights of the crop glands in different birds, and also that the increase in the mean weight is relatively small as the dose increases. In later experiments injections were given twice daily for 7 days; these were performed in April and May and the increase in the weight of the crop glands was much greater; the mean results of two experiments both carried out with the same extract (a stronger extract than that used for Table XLIX) are given in Table L.

TABLE L. *Each figure is the mean weight (gm.) of glands from five birds.*

Daily dose of extract	*Nil*	*0·025 c.c.*	*0·05 c.c.*	*0·1 c.c.*
Experiment 1 .	1·2	1·35	4·1	10·1
,, 2 .	1·15	2·1	4·5	9·8

Comparison of two preparations. A comparison of two preparations is made by taking four groups of ten pigeons; the weights of the different birds should be as similar as possible, and the total weight in each group should be about the same. Two of the groups receive doses of the standard, the dose given to one group being twice the dose given to the other group. Similarly, two of the groups receive doses of the unknown preparation. Doses are injected twice a day for 7 days. On the eighth day the mean weight of the crop glands in each group is determined.

III. The Growth Hormone

The estimation of the growth hormone present in anterior lobe extract is usually made in rats from which the pituitary gland has been removed.

Hypophysectomy in the rat. The following is Bülbring's description; it is a modification of the technique used by Selye. The rat is anaesthetized with ether. The wooden board to which the rat is fixed for the operation is 26 cm. long and 20 cm. wide (see Fig. 44). In the middle line, 5 cm. from one end, a metal hoop is driven into the wood, so that the upper incisor teeth of the rat will just go underneath. Three cm. farther a head-holder is fixed, made out of thick wire, which grips the back of the head and is just sufficiently pliable to be adapted to each animal. The legs are drawn backwards and fixed with wide rubber bands which are pulled through holes in the wood and fastened underneath. It is important that the rat should lie quite straight on its back and that the head cannot turn sideways. The field of operation is moistened with alcohol, and a median skin incision is made 2 cm. forwards from the sternum. The large salivary glands are pushed to the side, the muscles over the trachea are divided in the midline, and slightly below the thyroid a small hole is made in the trachea with a needle; a fine glass capillary is inserted through the hole. A short glass tube bent in an angle of about 135° and drawn out to a capillary is most suitable for this purpose. The capillary lies in the line of the trachea, but the thicker part of the tube falls on the right side of the rat, so that it is not in the way of the operation; by putting a piece of cotton-wool moistened with ether under the end of the tube, more anaesthetic can be given.

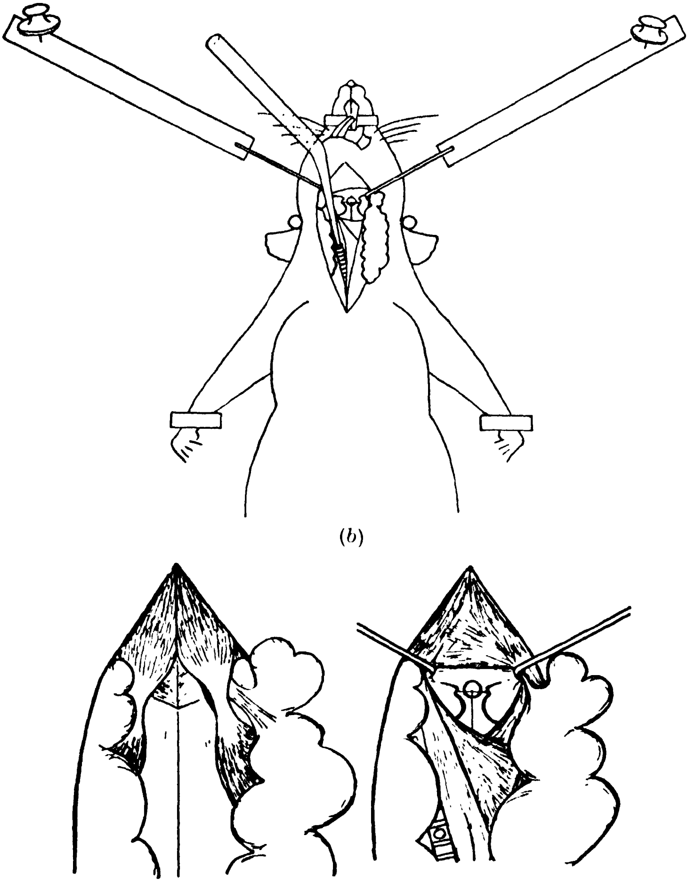


FIG. 44. Preparation for hypophysectomy. (*a*) Exposure of neck muscles, with left salivary gland everted. The black mark medial to the tendon of the digastric indicates the spot for the deep dissection. (*b*) Shows the method for fixing the rat, the position of the trachea tube, and the exposure of the base of the skull. (*c*) Enlargement to show point where the drill is applied. See text.

A deep blunt dissection is first made directly medial to the tendon of the left digastric muscle. The round mastoid process which stands out prominently in the rat is felt immediately, and the operator works his way medial to the large lingual nerve to the base of the skull, slowly approaching the midline. The nerve and the large vessels are held to the left side of the rat with a small blunt hook pulled by a rubber band, and the

pharynx and trachea are drawn to the right side by a second hook. By rubbing with small balls of cotton-wool, the base of the skull is exposed clean; in doing this care must be taken not to injure the vessels medial to the mastoid process leading to the inner ear, or the posterior wall of the pharynx.

The following picture should then present itself. Right and left are the globular mastoid processes between which runs transversely a bluish suture, the spheno-occipital suture. From behind, a bony ridge runs longitudinally up to the middle of this suture; this ridge can often be felt more easily than seen. The point where the hole is to be made through the skull with the dental drill lies just in front of the junction of the longitudinal ridge and the bluish suture. The drill should be No. 10, 11, or 12, according to the size of the rat. If the hole is made farther forward the venous sinus is easily opened, and if it is made farther back the hinder portion only of the pituitary is exposed, and the whole gland cannot be removed with certainty. The same is true if the hole is made to the side. In penetrating the skull with the drill, the last very thin layer of bone is preserved and then removed with a fine hook such as is used in ophthalmology. The hook is used to tear the membrane covering the rose-coloured pituitary. To suck out the gland, a glass tube attached to a vacuum pump is used; the mouth of the tube should pass easily through the hole in the skull. The sucking must be done carefully, otherwise bleeding from the sinus occurs. If the pituitary has been completely removed the bleeding soon stops by itself. If the operation has been successful the dura can be seen, and under it the brain, which pulsates with each respiration of the rat. A small ball of cotton-wool is left for a short time in the hole in the skull. Meanwhile, hooks and trachea cannula are removed, and if necessary the trachea is freed from mucus by the pump. After taking out the cotton-wool the skin is sewn together and the rat breathes naturally. The mortality depends on the amount of practice the operator has had, and on the duration of the operation. Without assistance the operation as described here takes 15 minutes, including the induction of anaesthesia. A binocular microscope, magnifying about 10 diameters, and a good lamp attached beneath it, are a great help for the quick and certain removal of the pituitary.

Performance of the test. For the test rats are hypophysectomized at a weight of 80–100 gm. They are then arranged in groups of 6–10 rats. To compare an unknown preparation with a preparation used as standard, four groups are necessary of which two receive the standard and two receive the unknown. The dose of the standard given to one group is twice the dose given to the other group; the dose of unknown given to one group is twice the dose given to the other, &c. The rats must be given a uniform diet on which normal rats will grow, and bread and milk is added to this. If the rats are kept on this diet for 1 week after the operation, it can be verified that growth has been arrested. The doses are given daily for 7 days by intraperitoneal injection, and the rats are weighed at the beginning and the end of this time. The mean weight change in each group is then determined.

To calculate the relative potency of the unknown preparation the mean results for all four doses are calculated and the relative potency is determined as on p. 184.

An example of the effect of different daily doses of an extract is given in Table LI. The experiment was carried out by Bülbring (1938). See also Fig. 45.

TABLE LI. *Weight change (gm.) per rat in seven days.*

Dose	*Nil*	*0·0083 c.c.*	*0·0166 c.c.*	*0·0332 c.c.*
	−2	+5	+8	+17
	−7	+16	+11	+19
	+5	+11	+11	+22
	+1	−6	+4	+12
	+5	+8	+14	+10
	−4	+8	+9	+10
Mean . .	0	+7	+9·5	+15

Adrenocorticotrophic hormone. A method of estimating this hormone was described by Sayers and Sayers (1946). It depends on the observation that the administration of the hormone causes a depletion of ascorbic acid in the suprarenal glands. The authors use hypophysectomized rats; they have observed that the amount of ascorbic acid is the same in the two adrenal glands, and therefore the left adrenal gland is first removed in order to determine the amount present.

Male rats are used weighing between 120 and 160 gm. The test is carried out 24 hours after the operation for the removal of the hypophysis. Each rat is anaesthetized with sodium

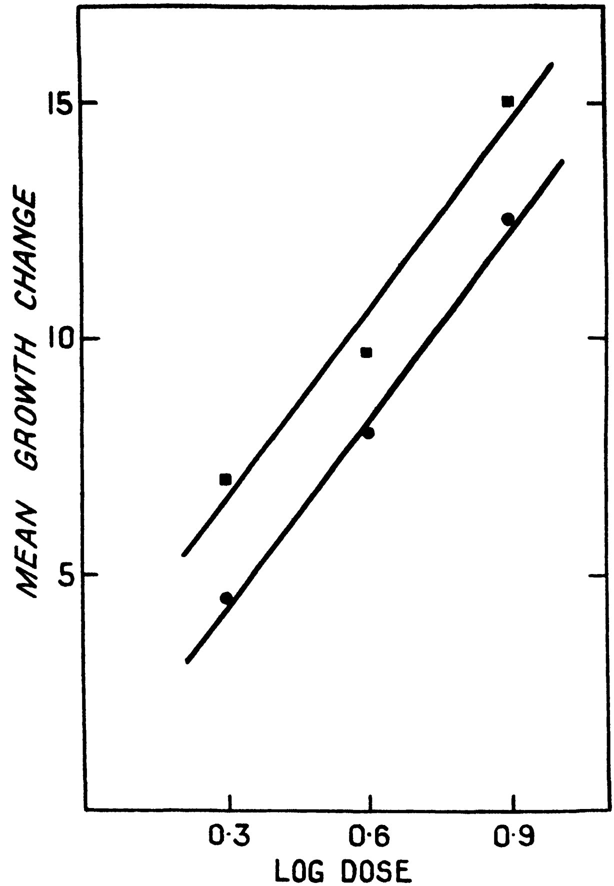


FIG. 45. The lines show the relation between the log dose of anterior lobe extract and the mean growth change in hypophysectomized rats in 7 days for two different extracts. (Bülbring.)

pentobarbitone (4 mg. per 100 gm.) injected intra-peritoneally, and the left suprarenal is removed. The solution to be tested is then given by injection into a tail vein. One hour later the right suprarenal is removed. The response to the injection is expressed as the difference between the concentrations of ascorbic acid in the right and left adrenal glands.

The excised adrenal is dissected free from fat on a filter-paper. It is then weighed, care being taken to avoid loss of moisture during dissection and weighing. The gland is placed in a 15 c.c. conical centrifuge tube containing 12 c.c. of 4 per cent. trichloracetic acid and a few grains of sand. A glass rod is used to crush and grind the tissue. Norit (0·3 gm.) is added and mixed with the contents of the tube, and analyses are made on aliquot parts of the filtrate by the method of Roe and Kuether (1943). The concentration of adrenal ascorbic acid is 300–500 mg. per 100 gm. tissue, and it is reduced by 20–190 mg. per 100 gm. according to illustrative figures given by Sayers, Sayers, and Woodbury (1948).

Gonadotrophic Hormones

Follicle-stimulating substances. These substances can be estimated by the method of Levin and Tyndale (1937) on the uterus of the mouse. They use mice of an in-bred strain which weigh from 7 to 10 gm. on the twenty-first to twenty-third day of life. The material to be tested is injected once daily on 3 consecutive days and the mice are killed on the fourth day. When the mice are killed, the uterus is removed, stripped of its mesentery, freed of intra-uterine fluid by pressure against a piece of dry filter-paper, and is weighed. The whole process is complete in 2 minutes for each animal. The authors found that the mean weight of the uterus was $5{\cdot}9 \pm 0{\cdot}3$ mg. for control mice, and that it rose according to the amount of FSH injected up to $44{\cdot}7 \pm 1{\cdot}6$ mg.

Luteinizing substances. These substances have been referred to by Greep, van Dyke, and Chow (1942) as interstitial cell-stimulating hormone or ICSH, giving the hormone itself the name metakentrin. ICSH can be assayed in hypophysectomized rats by its action in promoting the release of androgens which are most easily measured by the effect on the weight of the ventral prostate of the rat.

Chorionic Gonadotrophin

International standards. Standards exist for Chorionic Gonadotrophin and for serum Gonadotrophin. The unit of the former is the activity present in 0·1 mg.; the unit of the latter is the activity present in 0·25 mg.

Method using the rat's prostate. The method was originally devised by Korenchevsky, Dennison, and Simpson (1935) and will be described as performed by Dr. J. A. Loraine (1950) in the Department of Pharmacology at Edinburgh, to whom I am indebted for the details supplied. Some workers have used the rat's seminal vesicles (Watts and Adair, 1943), but this is less accurate, and the error of the test using weights of prostates is extremely small. Pregnancy urine can be assayed directly without extraction, since it contains enough hormone not to require concentration and the oestrogens present do not interfere with the test. Dorfman, Dorfman, and Grim (1948) have used the testis of the chick.

Rats weighing from 30 to 45 gm. are suitable for the test, the weight being more important than the age. Actually Loraine has used rats from 19 to 23 days old from a Wistar colony. It is not important to divide letters evenly among the groups.

The rats are injected once daily for 3 days. The standard is given to two groups, the doses for these groups being 2 units and 4 units respectively. Each sample of pregnancy urine is given to one group, the dose being a total of 0·5 c.c. urine given in the three injections. This dose is too large in normal pregnancy during the first 100 days, and also in a proportion of diabetic women (who are probably pre-eclamptic) throughout pregnancy. The urine from these patients should be given in a total dose of 0·3 c.c., or even 0·15 c.c. The number of rats required in a group is five.

The rats are killed on the fourth day, the prostate and seminal vesicles being carefully dissected and fixed in Bouin's solution for 24 hours. On the fifth day the organs are freed from fat, the seminal vesicles are discarded, and the prostate is dried by pressing it between pieces of filter-paper. By dissection from below the lobes of the prostate are separated from the urethra, and the whole prostate, both ventral and posterolateral lobes (see Fig. 40), is weighed. The results are expressed as mg. prostate per 100-gm. rat.

The details of a comparison of pregnancy urine with the standard are given in Table LII, in which ten rats were used in each group. The mean weight of prostates of rats injected with standard is then plotted against log dose, and the dose of the

standard which would have had the same effect as the dose of the unknown urine is thus determined.

TABLE LII. *Weights of prostate mg. per 100 gm.*

	Group 1, 2 units	*Group 2, 4 units*	*Group 3, 0·5 c.c. urine*
	110	178	160
	110	180	160
	137	196	159
	148	163	156
	135	142	154
	152	202	161
	135	188	143
	140	148	146
	127	170	145
	140	162	134
Mean .	133	173	152

It is not essential to use the standard in every test, but to use the standard only once in two weeks.

REFERENCES

ARON, 1930, *Rev. Française d'Endocrin.* **8**, 472.
—— 1930, *Compt. Rend. Soc. Biol.* **105**, 974.
—— 1932, ibid. **109**, 218.
—— 1932, ibid. **110**, 716.
BÜLBRING, 1938, *Quart. J. Pharm. Pharmacol.* **11**, 26.
DORFMAN, DORFMAN, and GRIM, 1948, *Endocrinol.* **42**, 93.
GREEP, VAN DYKE, and CHOW, 1942, ibid. **30**, 635.
HEYL, 1933, *Acta Brev. Neer.* **3**, 111.
KORENCHEVSKY, DENNISON, and SIMPSON, 1935, *Biochem. J.* **29**, 2522.
LEVIN and TYNDALE, 1937, *Endocrinol.* **21**, 619.
LOESER, 1932, *Arch. exp. Path. Pharmak.* **166**, 693.
LORAINE, 1950, *J. Endocrinol.* **6**, 319.
RIDDLE, BATES, and DYKSHORN, 1933, *Amer. J. Physiol.* **105**, 191.
ROE and KUETHER, 1943, *J. Biol. Chem.* **147**, 399.
ROWLANDS and PARKES, 1934, *Biochem. J.* **28**, 1829.
SAYERS and SAYERS, 1946, *Fed. Proc.* **5**, 200.
SAYERS, SAYERS, and WOODBURY, 1948, *Endocrinol.* **42**, 379.
SEVRINGHAUS, 1933, *Z. Zell. Mikro. Anat.* **19**, 653.
WATTS and ADAIR, 1943, *Amer. J. Obstet. and Gyn.* **46**, 183.

CHAPTER XII

VITAMIN D

The international standard. At a conference which met in London in 1931 it was recommended that the international standard for vitamin D should be a solution in olive oil of irradiated ergosterol prepared in the National Institute for Medical Research, London. At a second conference in 1934 it was recommended that this solution should be retained as the international standard, and that when it was exhausted it should be replaced by an equivalent solution of pure crystalline vitamin D in olive oil containing 0·025 μg. (0·025γ) in 1 mg. of solution. Crystalline vitamin D is calciferol or vitamin D_2, $C_{28}H_{43}OH$.

The standard solution is distributed in bottles of brown glass which are full; they contain about 10 c.c.; the bottles are stored at a temperature below 0° C. in the National Institute for Medical Research. The standard solution should always be kept at a temperature below 0° C.; when dilutions are to be prepared the solution should be allowed to remain at room temperature long enough for it to liquefy completely.

The unit. The international unit is defined as the vitamin D activity of 1 mg. of the international standard solution of irradiated ergosterol, which has been found equal to that of 0·025 μg. (0·025γ) of crystalline vitamin D. The international conference considered that comparisons between unknown preparations and the international standard should always be made on rats, and that if they were made on other species the potency could not be expressed in international units.

Methods

Comparisons of preparations containing vitamin D are made on the rat in three ways: by the line test, by X-rays, and by determination of the ash in the bone. Comparisons are made on the chicken in two ways: by determination of the ash in the bone and by X-rays.

I. The Line Test

(McCollum, Simmonds, Shipley, and Park, 1921)

Preparation of the rats. If young rats are given a rachitogenic diet they do not always develop rickets. To ensure that rickets will develop it is necessary to breed a colony for several generations in a room receiving very little daylight, and on a uniform diet. The following is a diet which Coward has used successfully (Coward, Cambden, and Lee, 1932):

Yellow maize, ground	65 gm.
Whole wheat, ground	20 ,,
Dried milk	20 ,,
Casein	9 ,,
Dried yeast	5 ,,
Sodium chloride	0·5 ,,
Calcium carbonate	0·5 ,,

Young rats are taken from the colony at a weight of 50–60 gm. and given the following rachitogenic diet for 3 weeks:

Yellow corn ground	76 per cent.
Gluten flour	20 ,,
Calcium carbonate	3 ,,
Sodium chloride	1 ,,

The ratio of calcium to phosphorus in this diet is 4 to 1.

The rats are weighed weekly, and those rats which do not grow are not used. At the end of the 3 weeks' preparatory period the test period begins, which is usually 10 days, but is shortened by some workers to 5 days. During the test period the rats are given the same diet. In experiments described by Dyer (1931) the average weight of rats at the beginning of the preparatory period was 54·5 gm., at the beginning of the test period it was 73 gm., and at the end it was 81·1 gm.

The response of the rats. During the test period the effect of vitamin D, administered either daily or as a single dose, is being exerted. The effect is observed by killing the rats and cutting off their forelegs 1 cm. behind the wrist. The distal ends of the radii and ulnae, freed from adhering tissues, are placed in a 4 per cent. solution of formaldehyde. After 24 hours the bones are washed in distilled water, and the radius and ulna of one leg are separated. Each bone is cut in two longitudinally in a plane parallel to that of the cut which divided the bones. The

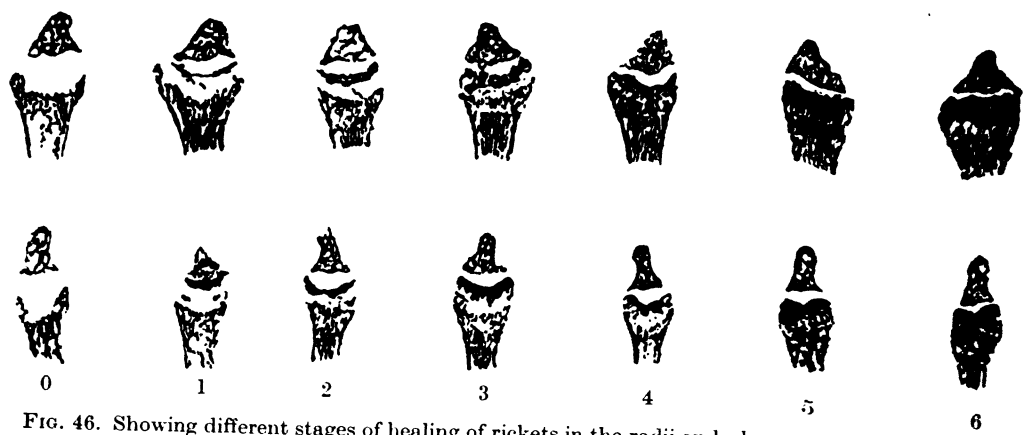


Fig. 46. Showing different stages of healing of rickets in the radii and ulnae of rats. 0 is severe rickets, 6 is fully healed rickets. (Dyer, *Quart. J. Pharm. Pharmacol.* 1931, **4**, 503.)

cut portions then stand in 1·5 per cent. silver nitrate, with cut surfaces uppermost, for about 3 minutes and are transferred to small porcelain dishes containing distilled water. Exposure to light causes a darkening of the bones due to a deposition of colloidal silver on the calcified areas.

The bones so prepared have the appearance shown in Fig. 46. On the left are bones from completely rachitic rats which have received no vitamin D during the test period; these have a wide

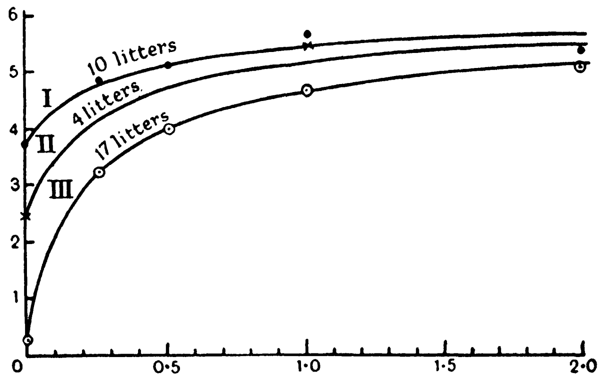


FIG. 47. Curves relating dose of standard vitamin D to the mean degree of healing of rickets (ordinate). See text. (Key and Morgan, *Biochem. J.* 1932, **26**, 196.)

uncalcified region at the epiphysis. On the right are bones from rats which have received a large dose of vitamin D and in which calcification is normal. The bones in the upper row of Fig. 46 are radii, and those in the lower row are the corresponding ulnae. The seven pairs represent the different stages which can be distinguished between the two extremes, and for the purpose of describing any bone the numbers 0–6 are given to them, as shown in the figure. Dyer, and later Key and Morgan (1932), studied the mean effect in groups of rats of different doses of irradiated ergosterol, using the preparation which in 1931 was the British standard. The effect in the bones of a rat was expressed by comparing the bones with Fig. 46 and taking the number of the most similar bone as an estimate of the healing. The results are shown in Fig. 47 which relates the dose of vitamin D to the mean degree of healing. Three curves were obtained; curve III included all the rats with severe rickets

(from 17 litters), curve II included 4 litters with moderate rickets, while curve I was drawn from the results of 10 litters with slight rickets. Unless the rats have severe rickets, the flatness of curve I shows that the error of the method is very high.

Dosage of the rats. During the test the vitamin D can be given daily, or more conveniently as a single dose at the beginning of the test. Coward and Key (1934) showed that the same mean degree of healing was obtained both for irradiated ergosterol and for cod-liver oil when a daily dose was given for 10 days as when the same total dose was given on the first day. The total dose of the standard is about 5 units, and the total dose of a sample of cod-liver oil is about 50 mg., depending on the potency of the oil. Since the unit is the activity present in 1 mg. of the international standard, 5 units are contained in 5 mg. The dose given to each rat is not less than 1 drop weighing 20 mg. (delivered from a pipette calibrated to deliver a drop of this size), hence the standard solution must be diluted four times with olive oil. The dilution is made by weighing 1 gm. of the standard and adding 3 gm. of olive oil. Each rat of the group receiving the standard must have a litter mate in the group receiving the unknown sample; no attention is paid to the sex of the rats.

Example of the method. The results of a comparison of a sample of cod-liver oil with the standard are given in Table LIII. Rats from four litters were used, those of each litter being evenly divided between the cod-liver oil and the standard. The figures for the expression of healing were obtained by comparing the bones with those in Fig. 46. The average healing produced by the cod-liver oil was 2·1, and that by the standard was 2·25. The curve III in Fig. 47 was then consulted, from which it was seen that healing in the ratio of 2·1 to 2·25 is produced by doses in the ratio of 0·146 to 0·161.

$$\text{Hence 50 mg. cod-liver oil} = \frac{0{\cdot}146}{0{\cdot}161} \times 5 \text{ units.}$$

Therefore 1 gm. cod-liver oil contains 90 units.

Experiment has shown that the slope of the curve varies according to the severity of the rickets, and the curve should not be used if the average healing in the two groups is very different. The experiment should then be repeated choosing

other doses. The use of the curve can be avoided by taking three groups of rats instead of only two. Two groups are given doses of the standard, the one dose being twice the other. The third group receives the unknown sample, and the dose of this is chosen in the hope that the healing produced will be intermediate between the healing produced by the two doses of the standard. In calculating the result the healing is assumed to be directly proportional to the logarithm of the dose; the calculation is most easily made by constructing a graph.

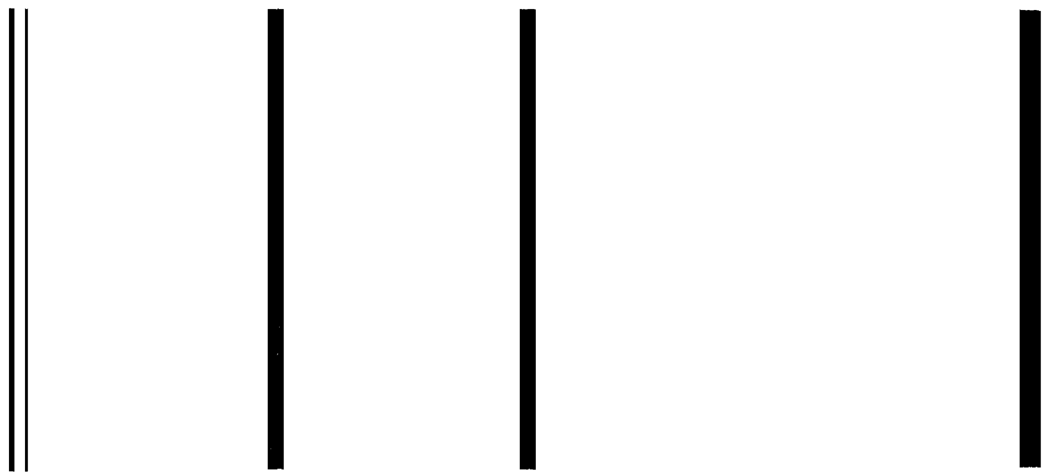
TABLE LIII

Dose	*Litter*	*Numerical expression of healing in each rat*	*Average healing*
50 mg. cod-liver oil	A	3·0	2·10
,, ,, ,,	A	2·5	
,, ,, ,,	A	2·0	
,, ,, ,,	B	1·0	
,, ,, ,,	B	1·5	
,, ,, ,,	C	2·0	
,, ,, ,,	C	3·5	
,, ,, ,,	D	2·5	
,, ,, ,,	D	1·5	
,, ,, ,,	D	1·5	
5 units standard	A	3·0	2·25
,, ,,	A	3·0	
,, ,,	A	2·5	
,, ,,	B	1·5	
,, ,,	B	1·5	
,, ,,	C	2·0	
,, ,,	C	3·5	
,, ,,	D	1·0	
,, ,,	D	2·5	
,, ,,	D	2·0	

II. The X-ray Method

The X-ray method is essentially similar to the line test, the chief difference being that the effect of the vitamin D which is given to the rats is observed by X-ray examination of the bones. The use of X-rays makes it possible to examine the rats at the end of the preparatory period to see that each rat is fully rachitic before it is used for the test. In a long-established colony of

2 3 4 6



7 8 9 10 11 12

FIG. 48 X-ray pho phs o bon of rats sho g the tage of healing of rickets. Bou Bruce, F d Web Med. R Coun. *Spec. Rep. Ser.*, 1931, no 8.)

rats severe rickets at the end of the preparation period is usually seen in every rat, but in colonies which are not more than 3 or 4 years old the development of rickets is sometimes irregular; in selecting rats for the test from a recent colony the use of X-rays is therefore of great value in diminishing the error. Apart from this, while the X-ray method is more expensive, it is rather more accurate than the line test. A valuable monograph on the X-ray method has been written by Bourdillon, Bruce, Fischmann, and Webster (1931).

The calculation of the result. Bourdillon and his co-workers give a series of photographs, shown in Fig. 48, representing the stages seen at the end of the test period according to the extent to which the rickets is healed. Photograph No. 1 is that showing the smallest detectable healing, and No. 12 is that showing complete healing. The number given to the photograph is called the 'scale value'. Rats are prepared as in the line test and at the end of the preparatory period are given daily doses of vitamin D for 10 days. The X-ray photographs of their bones are then compared with the photographs in Fig. 48 and the scale value for each rat is recorded.

We may take as an example the comparison of a sample of cod-liver oil with the standard, in which the doses of cod-liver oil were chosen on the assumption that the cod-liver oil contained 100 units per gm. In practice 10 rats should be used for the standard and 10 rats for the unknown sample, but in this example 4 rats are taken for each. It is not necessary to give the same dose to all rats receiving the standard; each rat can have a different dose. But for each rat receiving a dose of the standard a rat of the same litter receives a dose of the unknown which is expected to be equivalent. Thus, if a rat received 1 unit of the standard, then a rat of the same litter receives 0·01 gm. of the cod-liver oil. The results are shown in Table LIV. The difference in the scale-value for each pair of rats is observed and the mean difference calculated. Now Bourdillon *et al.* observed that an increase in scale-value of 2 was caused by doubling the dose; that is to say, an increase in scale-value of 2 corresponded to an increase in dose of log 2 which is 0·3010. Therefore, a difference in scale-value of $-1{\cdot}25$ corresponds to a decrease in dose of $\frac{0{\cdot}3010 \times -1{\cdot}25}{2} = -0{\cdot}1881 = \bar{1}{\cdot}8119$. This is the

TABLE LIV

Dose of standard (units)	*Dose of cod-liver oil mg.*	*Scale-value for standard*	*Scale-value for cod-liver oil*	*Difference*
1·5	15·0	8·0	6·5	−1·5
0·75	7·5	6·5	4·5	−2
0·375	3·75	3·5	2·0	−1·5
0·187	1·87	1·0	1·0	0
			Sum	−5
			Mean	−1·25

logarithm of 0·6485. Therefore the cod-liver oil had 0·6485 of the expected potency; that is, it contained about 65 (64·85) international units of vitamin D per gm.

III. Estimation by the Ash Percentage

The estimation of vitamin D by the ash content of bones is based on the method first described by Chick, Korenchevsky, and Roscoe (1926), and later developed by Hume, Pickersgill, and Gaffikin (1932). Young rats of about 40 gm. weight are fed on the rachitogenic diet already described for 28 days, and during the whole of this time they receive the doses of vitamin D. The rats are then killed and their leg-bones are extracted to remove fat; the percentage of ash in the fat-free bone is determined.

Three groups of ten rats are required, and the rats of each litter are divided as evenly as possible among the groups. Two of the groups receive doses of the standard, and the third group receives a dose of the unknown sample. All the animals in one group receive the same dose. The daily doses of the standard should be 0·1 unit and 0·2 unit for the respective groups, and the daily dose of the unknown sample should be a dose expected to be equal to 0·15 unit.

The bones taken for examination are either the femur, tibia, and fibula of the hind legs, or else the humerus and femur. They are cleaned, broken across, dried, and then extracted with alcohol and ether to remove the fat. The extraction is carried out in a Soxhlet apparatus. The extracted bones are weighed when dry, and then heated in a crucible until ash alone remains

and further heating does not reduce the weight. The final heating is completed in an electric muffle furnace. The result for each animal is expressed as the percentage of ash in the dry extracted bone.

Although there is a considerable variation in the percentage of ash in the different rats in a group, the constancy of the mean percentage in ten rats is illustrated by the results given in Table VI (p. 18). The mean percentages for the three groups are calculated, and the percentages for the groups receiving the standard are plotted as ordinates against the logarithms of the doses as abscissae. A line is drawn through the two points, and the mean ash percentage in the group receiving the unknown sample is marked on this line. The abscissa of this point is the logarithm of the dose of the standard to which the dose of the unknown sample is equivalent.

Estimation of vitamin D in chickens. Evidence has been obtained by Massengale and Nussmeier (1930) that the relative potency of cod-liver oil, and of irradiated ergosterol determined on rats, differs from the relative potency determined on chickens; these workers have found that cod-liver oil is relatively much more effective in chickens. The conclusion has been drawn that the vitamin present in cod-liver oil differs from the active substance in irradiated ergosterol. Evidence of a similar kind has recently been obtained by Bills that the vitamin D of tunny-fish oil differs from that of cod-liver oil. The validity of this evidence depends partly on the accuracy with which comparisons can be made in chickens, and hitherto no methods as good as those in which rats are used have been described. Two of the methods in which chickens are used are the method of Massengale and Bills (1936) in which the percentage of ash in the bone is estimated, and the method of Grab (1936) in which X-ray photographs are taken.

IV. Estimation of Ash Percentage

Preparation of the chicks. If it is possible the chicks should be obtained from a colony of one breed, such as single-comb White Leghorns, fed on a diet of which 100 gm. contains 200 international units of vitamin D as cod-livei oil. Newly hatched chicks are put in pens which are heated and illuminated with red light; they are given no food during the first 48 hours. From

the third to the seventeenth day they are fed on the following diet:

Ground yellow maize	56 per cent.
Wheat bran	10 ,,
Linseed oil meal	10 ,,
Wheat gluten	10 ,,
Dried skimmed milk	9 ,,
Calcium carbonate	2 ,,
Sodium chloride	1 ,,
Vegetable oil	2 ,,

The vegetable oil is maize oil or cotton-seed oil which is used in the test period as a vehicle for the cod-liver oil to be tested. The calcium-phosphorus ratio of this diet is 2:1, but the ratio is unimportant since chickens develop rickets unless vitamin D is supplied, whatever the ratio. The preparation recommended by Massengale and Bills can be summarized by saying that chicks are taken from hens which have been richly supplied with vitamin D, and are fed on a diet free from D for the first 2 weeks.

Performance of the test. The test period is 28 days, in which the chicks are fed on the same diet together with the cod-liver oil or irradiated ergosterol. The chicks are kept five in a cage, in a dim light, at an even temperature of 27° C. As already stated the vitamin D is given as a constituent of the diet; in this respect the method could be improved, for most workers believe that it is better to give the dose of vitamin D by mouth to each bird rather than to mix the dose in the food. At the end of the 28 days the chickens are killed, and one femur is taken for determination of the ash percentage. The femur is dissected immediately after death; it is placed in boiling water for 1 minute and scraped free of adherent tissue, taking care not to remove the cartilaginous epiphyses. The bones are broken in three pieces, extracted in a Soxhlet apparatus for 12 hours with alcohol, and 12 hours with ether. The extracted bones are dried to constant weight at 100° C., and are then heated in an electric muffle furnace to reduce them to ash. The ash is weighed, and the percentage of ash in the dry and fat-free femur is calculated.

In comparing the relative potency of two substances three groups of chicks should be used, each group containing 10 chicks. (In view of the very large increase of dose necessary to give a reasonable increase in ash percentage, it is wiser to use 20 or

30 chicks in one group; Massengale and Bills, however, say 10 chicks.) Two of the groups are given the standard, the dose for the chicks in one group being twice the dose for the chicks in the other group; the third group is given the unknown sample. The result is calculated in the same way as in the ash percentage method in rats.

Massengale and Bills have constructed a curve relating the percentage of ash to the dose of cod-liver oil, and also to the dose of irradiated ergosterol. They state the amount of cod-liver oil in international units determined by the line test on rats. Actually the cod-liver oil was found to contain 77 units per gm., hence a dose stated to be 9·5 units of cod-liver oil was 0·123 gm. per 100-gm. diet. The chief results are given in Table LV. The average number of chicks used to determine each

TABLE LV

(From Massengale and Bills, Fig. 1, 1936)

Dose of cod-liver oil (units per 100-gm. diet)	*Mean ash percentage*	*Dose of irradiated ergosterol (units per 100-gm. diet)*	*Mean ash percentage*
3	37·0	150	38·9
9·5	42·75	300	41·4
19·5	46·75	650	44·8
38	48·3	1,250	46·0

figure for the mean ash percentage in Table LV was 37. The figures illustrate the very slight effect on the ash percentage produced by a large increase in dose; thus an increase from 3 units to 9·5 units of cod-liver oil, an increase, that is to say, of over 300 per cent., increased the ash content from 37·0 to 42·75. Massengale and Bills think that the relation between dose and effect is fairly constant, and can be assumed to apply to ordinary tests. Since Hume and her colleagues (1932) found that this was untrue for rats, it is safer to determine the relation in each test by taking two doses of the preparation used as standard.

V. The X-ray Method

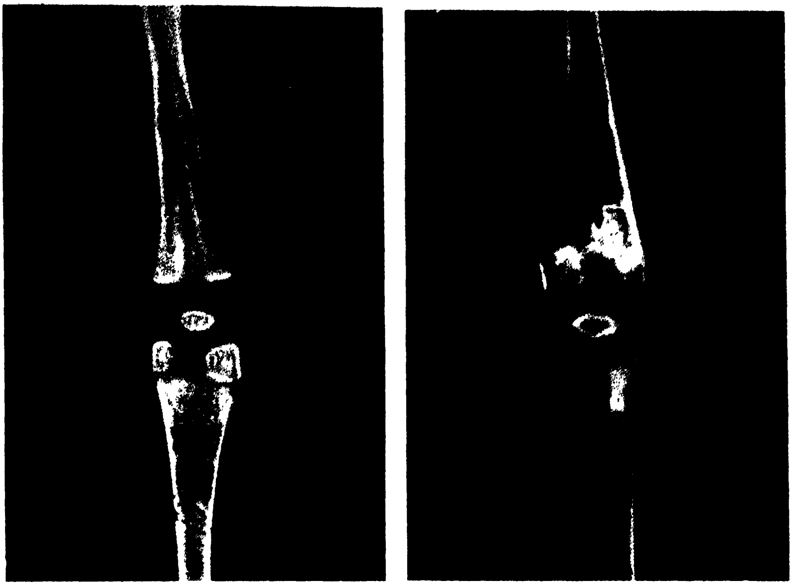
Preparation of the chicks. The preparation of chicks recommended by Grab (1936) is similar to that described above. Care is taken to see that the hens from which the eggs are taken

have plenty of vitamin D in their food. During the first 3–5 days of life of the chicks they are fed on slices of hard-boiled egg; they are then given a vitamin D-free diet of the following composition:

Ground maize	58 per cent.
Ground wheat	24 ,,
Extracted casein	12 ,,
Calcium carbonate	3 ,,
Calcium phosphate	1 ,,
Vitamin D-free yeast extract . .	1 ,,
0·05 per cent. solution of carotene in sesame oil	1 ,,

Since milk and yeast contain very little vitamin D, it is doubtful whether it is necessary to take trouble to extract them as suggested by Grab. Similarly, the maize should contain sufficient carotene without giving a solution of carotene in sesame oil. However, Grab finds that his chicks eat readily about 30 gm. daily. The ratio of calcium to phosphorus in this diet is 2·5 to 1. During the period in which they eat this diet the chicks are placed in separate wire cages, and the food is in tins outside the cages. Grab kept chicks on this diet for 8 weeks in order that the rickets should be severe.

Performance of test. During the test the doses of vitamin D are administered daily by stomach tube, for 14–17 days. Each dose is given to a group of ten chicks. At the end of the test X-ray photographs of the knee-joints are taken, and from this it is decided whether each chicken is suffering from rickets or not. Photographs taken from Grab's paper are reproduced in Fig. 49. Grab also took into account the signs of rickets in each chick before it was killed, and the macroscopic evidence of rickets obtained *post mortem*. He used a series of doses in the ratio of 1:4:16:64:256; thus the daily doses of crystalline vitamin D_2 (prepared from irradiated ergosterol) were $0{\cdot}2\gamma$, $0{\cdot}8\gamma$, $3{\cdot}2\gamma$, $12{\cdot}8\gamma$, and $51{\cdot}2\gamma$. Even the dose of $51{\cdot}2\gamma$ did not heal rickets in all chicks, while the dose of $0{\cdot}8\gamma$ healed rickets in some. These doses are in the ratio of 64:1, which means that unless large groups of chicks are used the error in estimating the potency of a preparation will be very great (see p. 24). Grab expressed the activity of each preparation in terms of the dose which cured rickets in eight out of ten chicks; he compared this dose with the



(a) (b)

FIG. 49. X-ray photographs of bones of chickens. (a) Severe rickets; (b) rickets healed. (Grab, *Zeitschr. f. physiol. Chem.* 1936, **243**, 63.)

minimal curative dose for rats. For crystalline vitamin D_2, the ratio of the rat dose to the chicken dose was 1:32 in one experiment in summer, and 1:100–400 in a second experiment in winter.

REFERENCES

BOURDILLON, BRUCE, FISCHMANN, and WEBSTER, 1931, Med. Res. Coun., *Spec. Rep. Ser.*, no. 158.

CHICK, KORENCHEVSKY, and ROSCOE, 1926, *Biochem. J.* **20**, 622.

COWARD, CAMBDEN, and LEE, 1932, ibid. **26**, 679.

COWARD and KEY, 1934, ibid. **28**, 870.

DYER, 1931, *Quart. J. Pharm. Pharmacol.* **4**, 503.

GRAB, 1936, *Zeitschr. f. physiol. Chem.* **243**, 63.

HUME, PICKERSGILL, and GAFFIKIN, 1932, *Biochem. J.* **26**, 488.

KEY and MORGAN, 1932, ibid. **26**, 196.

MASSENGALE and BILLS, 1936, *J. Nutrit.* **12**, 429.

MASSENGALE and NUSSMEIER, 1930, *J. Biol. Chem.* **87**, 423.

McCOLLUM, SIMMONDS, SHIPLEY, and PARK, 1921, ibid. **47**, 507.

CHAPTER XIII

DIGITALIS, STROPHANTHUS, AND SQUILL

I. Digitalis

The standard preparation. Since the digitalis standard was the first standard to be introduced for a substance other than a serum, its history is of some interest. The proposal to prepare this standard was put forward at a conference in Edinburgh in 1923 by Magnus, and the first international standard of 1926 was prepared by Magnus as the outcome of this proposal. This standard was a mixture of ten different samples of powdered digitalis leaf all dried at 55°–60° C., which was put into ampoules and sealed in Utrecht, and deposited for distribution in the National Institute for Medical Research, London. Since the mixture was made from ten different samples, its activity was presumed to be about equal to the mean activity of samples of digitalis; the standard was, therefore, available for anyone to compare with it an unknown sample of leaf to see how far this sample differed in activity from the mean activity of digitalis leaf.

The unit. At a conference in Frankfurt in 1928 a recommendation was made in the following terms: 'That, when the dosage of digitalis or its preparations is expressed in units of activity, the unit employed for any preparation and in any country should be an international unit which should be defined as the specific activity contained in 0·1 gramme of the international standard powder.'

The 1936 standard preparation. In 1935 the standard prepared by Magnus was nearly exhausted. The Frankfurt Conference had agreed that when this occurred, Bijlsma, the successor of Magnus in Utrecht, should be asked to prepare the new standard. In the meantime, however, a British standard had been prepared, and in 1935 the Permanent Commission on Biological Standardization, with the agreement of Bijlsma, decided to adopt the British standard as the new international standard, to be known as the international standard (1936).

The material used for the British standard consisted of nine different samples of the leaves of *Digitalis purpurea* obtained from four British and American growers. These samples were known to have been prepared by transfer of the freshly gathered leaves directly to drying-chambers at 55°–60° C. After mixing the samples, the mixture was dried in a special apparatus for 24 hours at 50°–58° C., and distributed into ampoules in a room supplied with warm, dry air. The filled ampoules were kept for 6 days in a low-pressure chamber over $CaCl_2$ at 30°–35° C. before being sealed; the final water content was found to be about 3 per cent. in accordance with a recommendation of the Frankfurt Conference.

Activity of the 1936 standard. The British standard thus prepared was compared with the international standard (1926) in three British laboratories, and in one in Holland, by the frog method and also by the cat method. The British standard was found to be the more active, and the results by the frog method gave it a higher relative potency than those by the cat method; whereas the potency by the frog method was 1·4 times that of the 1926 international standard, the potency by the cat method was 1·16 times.

Since this British standard is now adopted as the international standard, 1936, the potency relative to that of the international standard, 1926, is of great practical importance. The National Institute for Medical Research, London, recommends that the new standard be regarded as 1·25 times as active as the old, without respect to the method of comparison applied. If this recommendation is accepted, then 1 unit of activity, previously contained in 0·1 gm. of the 1926 standard, is present in 0·08 gm. of the international standard, 1936.

Methods of Comparison

I. The Frog Method

In the frog method an injection is made under the skin of the frog, and it is left for 1 hour, 3 hours, or overnight, according to the period preferred. In the 1-hour method the frogs are all killed at 1 hour after injection; those in which the heart is still beating are counted survivors. While the 1-hour method used to be official in the *United States Pharmacopoeia*, many

workers agree that the 12-hour (or overnight) period enables a sharper distinction to be drawn between the dead frogs and the survivors.

Preparation of solution for injection. The simplest method of obtaining the active principles in solution is to use a maceration process described by Berry and Davis (1935). The leaf is ground to a powder; of this, 5 gm. is put in a glass mortar, and 70 per cent. alcohol is added a few c.c. at a time until the total quantity is 50 c.c. The mixture is transferred to a wide-mouthed bottle of about 100 c.c. capacity; the bottle is closed, well shaken, and set aside over 2 nights. The mixture is then strained through fine muslin, and the tincture is filtered; 1 c.c. contains the activity of 0·1 gm. leaf.

Comparison of two tinctures. A tincture prepared from the standard can be compared with one prepared from an unknown leaf by any suitable method given in Chapters II and III. A comparison by Trevan's method will be described here. Tincture of digitalis is injected after being diluted with 0·6 per cent. saline; it is unnecessary to remove alcohol by evaporation. The dilution may be 1·5–10 c.c., or 4–10 c.c. according to the sensitiveness of the frogs and the potency of the tincture. Of this dilution 0·02 c.c. is injected per gm. Thus the dose in terms of undiluted tincture is from 0·003 to 0·008 c.c. per gm.

The frogs selected for any part of a test must all be of the same sex, so that when a certain number of males is injected with the unknown preparation, the same number of males is injected with the standard; but females may be used as well as males. Females swollen with eggs preparatory to spawning should not be used. The weights of frogs injected with the unknown preparation should correspond closely with the weights of frogs injected with the standard, since the dose is given proportional to the body-weight. Fromherz has shown that when doses are adjusted according to body-weight, heavier frogs appear more sensitive than lighter ones. It is, however, better to make the adjustment than not to make it, provided that the weights of frogs in the two groups are similar.

Observations are usually made on at least 2 days. On the morning of the first day 12 frogs are placed under glass jars in a uniformly lighted part of the laboratory. Each frog is weighed. Six frogs are then injected with the unknown preparation and

6 with the standard. For the injection, the needle of the syringe is inserted through the muscles of the thigh to bring the point of the needle under the skin covering the abdominal muscles. In 2 or 3 hours it is possible to see whether the doses are suitable, that is to say, whether about half of the frogs injected with the unknown are dying, and also about half of those injected with the standard. If so, a further group of 12 frogs is selected and injected as before. If, however, in either of the groups all of the frogs have died, or all appear unaffected, a different dose is chosen for the corresponding group of the second batch of 12 frogs. On the following morning further injections are made, and the test is continued until results have been obtained on 24 frogs injected with the standard, and on 24 injected with the unknown. The relation of the percentage mortality to potency according to Trevan's observations on *Rana temporaria* is given in Table LVI.

TABLE LVI

Mortality per cent.	*Potency*	*Mortality per cent.*	*Potency*
5	56	55	103
10	67	60	107
15	75	65	110
20	80	70	114
25	83·5	75	118
30	87	80	122
35	90	85	127
40	93·5	90	134
45	97	95	146
50	100	..	..

The following is an example of the examination of a tincture of digitalis.

During the first day of the test a dose of 0·00225 c.c. per gm. frog of the unknown tincture was administered to each of 12 frogs of which 9 died. A dose of 0·0035 c.c. of the standard tincture was given to each of 12 frogs of which 6 died. Hence the unknown tincture produced a mortality of 75 per cent. and the standard tincture a mortality of 50 per cent. The relative potencies corresponding to these mortalities are 118 and 100 respectively.

Hence

0·00225 c.c. unknown tincture

$$= \frac{118}{100} \times 0{\cdot}0035 \text{ c.c. standard tincture;}$$

or 1 c.c. unknown = 1·84 c.c. standard.

During the second day of the test a dose of 0·0025 c.c. per gm. of the unknown killed 5 out of 12, or 42 per cent. of frogs, while a dose of 0·004 c.c. of the standard killed 4 out of 12 frogs, or 33 per cent. These percentage mortalities correspond to potencies of 95 and 89 respectively.

Hence

$$0{\cdot}0025 \text{ c.c. unknown} = \frac{95}{89} \times 0{\cdot}004 \text{ c.c. standard;}$$

or 1 c.c. unknown = 1·71 c.c. standard.

The mean of the two results 1·84 and 1·71, which is 1·78, is taken as the true value. Hence the unknown tincture is 1·78 times as strong as the standard, and contains 1·78 units per c.c.

Trevan has calculated that the expected standard deviation of the result when 48 frogs are used for the test in the above manner is 10 per cent.

Chapman and Morrell's curve. Using frogs obtained in the neighbourhood of Ottawa, Chapman and Morrell (1931) have determined the characteristic curve for ouabain. They took many precautions, and their paper is in many respects a model for those interested in the construction of a characteristic curve. They observed less variation than Trevan, and their figures relating dose to percentage mortality are given in Table LVII. The characteristic curve is, therefore, not always the same in different countries.

II. Estimation by the Cat Method

The cat method was introduced by Hatcher and Brody in 1910 and was fully investigated in the laboratory of Magnus at Utrecht. The method, as described by de Lind van Wijngaarden (1926), is carried out in the following way. A cat is weighed; it should not be pregnant or lactating, and its weight should lie between 1·7 and 2·7 kg. It is anaesthetized with ether, and tracheotomy is performed so that artificial respiration

TABLE LVII

Mortality per cent.	*Potency*	*Mortality per cent.*	*Potency*
5	83	55	101
10	87·5	60	102·5
15	90	65	104
20	92	70	105·5
25	93	75	106·5
30	95	80	108
35	96	85	110
40	97·5	90	112·5
45	99	95	116
50	100	..	..

may be given with a mixture of air and ether vapour. This is done to ensure that death does not occur from respiratory failure. A cannula is inserted in the femoral vein for the slow infusion of the preparation of digitalis which is to be tested. The preparation to be given is put in a burette, from which it passes through a warming coil before reaching the vein. A tube passes through a stopper in the top of the burette, drawn out below the cork to a long fine capillary, and the solution leaves the burette at the rate at which air enters through the capillary. So long as the solution is above the bottom of the capillary, the effective pressure sending the solution into the vein remains constant. When the artificial respiration has begun, the time is noted and the intravenous infusion is started at a rate of about 1 c.c. per minute. The entry of the digitalis into the circulation slows the heart during the first 25 or 30 minutes, after which the pulse suddenly becomes very rapid, and in about 15 minutes more the heart is arrested in systolic contraction. This point is detected by palpating the chest, the intravenous infusion is stopped, and the amount of solution which has entered the vein is measured. The details of the test are as follows.

Preparation of the solution. Digitalis may be administered either as an infusion or as a diluted tincture. An infusion may be prepared by adding 250 c.c. tap-water to a large beaker containing 1·25 gm. powdered leaf. The beaker is put into a water-bath and the temperature inside the beaker is allowed to rise to 89°–91° C., where it is kept for 15 minutes with constant stirring. The infusion is filtered through muslin, and when it is

cool, sodium chloride is added to make 0·9 per cent., and the volume is adjusted once more to 250 c.c. Tincture of digitalis may be administered after diluting 10 c.c. to 200 c.c. with 0·9 per cent. saline.

The anaesthetic. Although ether has long been used as anaesthetic in this method, two disadvantages attend its use. The average lethal dose of a sample of digitalis leaf is smaller the greater the amount of ether; thus the lethal dose of a sample tested by de Lind van Wijngaarden in Utrecht was 90·5 mg. per kg., while, tested in the author's laboratory, in which more ether was used, it was 76·3 mg. per kg. As the infusion of digitalis proceeds, the amount of ether given should be reduced to a minimum. The second disadvantage of ether was first pointed out by Macdonald and Schlapp (1930), who showed that the variation in the figures for different cats tested with a given sample was much greater when ether was used than when the digitalis was administered to spinal cats. A series of papers has been published dealing with this question of anaesthetic (see Epstein, 1933, Chopra and Chowan, 1933, David and Rajamanickam, 1934, and Macdonald, 1934), in which all observers agree that the variation among cats is greater when ether is used, is less for paraldehyde, and still less for chloretone, chloralose, and urethane. Spinal cats, however, are best of all, and Macdonald has calculated that, to obtain a given degree of accuracy, twice as many cats are necessary if ether is used as if chloralose is used, and five times as many etherized cats as spinal cats. The different anaesthetics have been administered by the workers mentioned in the following manner. Paraldehyde was given by intraperitoneal injection of a 10 per cent. aqueous solution in a dose of 1·6–1·8 gm. per kg. Chloretone was given by intraperitoneal injection in a dose of 0·2 gm. per kg. dissolved in alcohol. Chloralose was given by intramuscular injection in a dose of 0·11 gm. per kg. dissolved in alcohol, and urethane was given by intramuscular injection in a dose of 1·8 gm. per kg. dissolved in hot water.

Spinal cats are prepared as described on p. 219; after section of the spinal cord they are left for a period of 4–5 hours so that the volatile anaesthetic may be removed; this causes delay in completing the test, but Macdonald argues that the smaller number of cats compensates for this.

The determination of the end-point. The end-point should be reached in not less than 30 and not more than 55 minutes; it may be more easily determined by taking a record of the blood-pressure than by palpating the chest, for when the heart stops in systolic contracture the blood-pressure falls to zero. This fall may be a sudden drop from a high point, or a gradual decline. Often the heart starts beating again after stopping once, and the point of final arrest must be observed.

Calculation of the result and number of animals required. In the method proposed by Hatcher the potency of digitalis leaf was expressed in cat units. The lethal dose of a series of cats was determined, and the results were expressed per kg. body-weight. The mean of these results was the cat unit. This figure is, however, not a satisfactory expression of the potency of a sample, since it varies according to the technique, being particularly affected by the anaesthetic used. Hence the cat method, like the frog method, must be compared with the standard leaf, since this comparison is unaffected by differences in technique.

There is no evidence that cats vary in sensitivity to digitalis at different times of the year, and it is, therefore, unnecessary to make simultaneous tests of the unknown and the standard. A determination of the average lethal dose per kg. cat of the standard leaf can be made at one time, and determinations of unknown samples of leaf can be made at later dates; the results of these determinations are compared with the result previously obtained for the standard. In order to obtain an expression of the potency of an unknown leaf in terms of the standard which may be expected to have a standard deviation of 10 per cent., 14 cats should be used for the standard and 6 cats for the unknown leaf. This calculation was made by Trevan on the basis of de Lind van Wijngaarden's findings, that the standard deviation of a single determination on a single cat is 13 per cent., de Lind van Wijngaarden used a minimum amount of ether as anaesthetic. From what has been said it will be realized that, for cats anaesthetized with other anaesthetics, or for spinal cats, the standard deviation is less, and a smaller number of cats will give the same accuracy. At present there are not sufficient data to say what the standard deviation is when other anaesthetics are used.

As an example of the method, figures taken from Macdonald's experiments (1934) on spinal cats are given in Table LVIII. A tincture of strophanthus was estimated in relation to the international standard ouabain. From the average lethal doses given in the table, it follows that 17·37/500 c.c. tincture = 0·098 mg. ouabain, and 1 c.c. tincture = 2·82 mg. ouabain.

TABLE LVIII. *Results obtained by Macdonald on spinal cats*

International standard ouabain. Lethal dose in mg. per kg.	*Tincture of strophanthus diluted 1/500. Lethal dose c.c. per kg.*
0·100	17·08
0·100	16·74
0·097	17·92
0·094	17·73
0·092	17·37
0·100	..
0·102	..
Average 0·098	17·37

III. Estimation by the Guinea-pig Method

This method, introduced by Knaffl-Lenz (1926), is in essentials similar to the cat method; it demands more skill in preparing the animal, but takes less time, and is more accurate since the variation among guinea-pigs is less than among cats, when these are anaesthetized with ether. Guinea-pigs weighing from 400 to 600 gm. are chosen and anaesthetized by the subcutaneous injection of urethane. A 25 per cent. solution in water is prepared and a volume corresponding to 1·75 gm. per kg. is injected. After 2–3 hours the animal is secured on a board and a little ether is given. The jugular vein from one side is then dissected in the neck and a vein cannula is inserted; as the vein lies surrounded by fat, this dissection requires some practice. Thereafter the ether is discontinued.

The diluted tincture or infusion of digitalis is then allowed to run slowly into the vein from a small burette, the strength of the solution corresponding to 1·25 per cent. of digitalis leaf or 0·125 unit per c.c., and the rate of infusion being about 0·5 c.c. per minute. During the infusion the heart is constantly palpated through the chest-wall; towards the end the beats are felt to be

feebler, and then the chest is opened so that the final cessation of the heart is observed directly. The period of infusion is usually from 15 to 20 minutes, the rate being reduced near the end. A series of guinea-pigs is tested with the standard, and the average figure for the lethal dose is determined; the average figure for the lethal dose of an unknown tincture can then be compared with the figure for the standard as in the cat method.

In Table LIX an example is given of the estimation of a sample of digitalin (*Digitalinum Purum Germanicum*) in terms of the international standard digitalis powder. The figures are taken from a paper by Gage (1933). The average figures show that 6·2 mg. of digitalin is equivalent to 1·55 c.c. of tincture prepared from the standard, which contains by definition 1·55 units. Hence the sample of digitalin contains 250 units per gm.

TABLE LIX

Digitalin. *Lethal doses in mg. per kg.*	*Standard tincture.* *Lethal doses in c.c. per kg.*	
5·3	1·18	1·33
5·6	1·46	1·43
6·5	1·55	1·74
7·1	1·56	1·40
6·1	1·35	1·85
6·6	1·81	1·59
..	1·63	1·78
Average 6·2	1·55	

IV. Estimation by the Pigeon Method

The pigeon method, introduced by Hanzlik (1929), depends on the observation that preparations of the cardiac glucosides, administered by intravenous injection, have an emetic action in pigeons. There is considerable variation in the sensitivity of different birds, as shown by Burn (1930), and it is necessary to determine the percentage of pigeons in which an emetic effect is produced by a dose of the unknown sample and also by a dose of the standard.

Pigeons weighing 250–450 gm. are used. Each bird is weighed, and while it is held by an assistant a few of the small feathers covering the underside of the wing at the point where the wing joins the body are plucked. The wing vein is then

visible under the skin. The injection is made with a sharp needle of fine bore; when the needle is withdrawn there is usually a haematoma under the skin, and there may be a little bleeding, but the haematoma will be slight if the needle is sufficiently fine. The less the haematoma the shorter the interval before the vein may be used again.

Digitalis is injected in the form of diluted tincture, the tincture being prepared as described on p. 296, and the dilution being made with 0·9 per cent. saline. The dose causing emesis in 50 per cent. of pigeons is usually in the range 0·15–0·4 c.c. of the undiluted tincture per kg., and it is convenient to dilute the tincture so that the volume injected is 0·1 c.c. per 100 gm. Thus a pigeon weighing 300 gm. receives 0·3 c.c.

To compare an unknown sample with the standard, the methods already described for frogs may be used (pp. 296–7). A characteristic curve may be determined, or Behren's method may be followed. Another method, first proposed by Krogh and Hemmingsen (1926), is to determine for each tincture a dose which causes emesis in less than 50 per cent. of birds, and also a dose causing emesis in more than 50 per cent. of birds; an example is given in Table LX.

TABLE LX

Tincture	*Dose, c.c. per kg.*	*No. of pigeons used*	*No. which vomited*	*Percentage response*
Standard	0·2	30	13	43
	0·3	30	22	73
Unknown	0·15	30	9	30
	0·2	30	17	57

The percentage responses for each tincture are then plotted on a chart as ordinates against the logarithms of the doses as abscissae. This is done in Fig. 50. A straight line is drawn through the two points for each tincture, and the abscissae corresponding to a 50 per cent. response are observed from the chart. The doses corresponding to these abscissae represent the relative potency of the two tinctures. In Fig. 50 the average emetic doses of the standard and unknown are 0·21 c.c. per kg. and 0·19 c.c. per kg. respectively, so that the unknown tincture

has 1·1 times the potency of the standard tincture, or contains 1·1 units per c.c. Probably the best way of making the comparison is to carry out the work on 2 days, using the same pigeons on each day. On the first day half the birds receive the standard and half receive the unknown. On the second day,

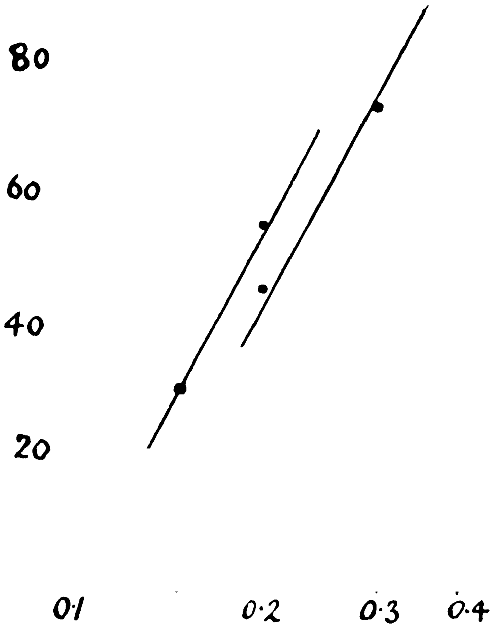


FIG. 50. Relation between the dose of tincture of digitalis (on logarithmic scale) and the percentage of pigeons in which vomiting occurred. The two lines are for two tinctures.

which may be 3 or 4 days after the first, the birds are injected by way of the opposite wing vein; the half which previously received the standard now receive the unknown, and vice versa.

II. STROPHANTHUS

(i) The Estimation of Tincture of Strophanthus

Standard preparation. The crystalline strophanthin which is obtained from *Strophanthus gratus*, and which is known as ouabain, was first proposed as a standard for tincture of strophanthus by the *United States Pharmacopoeia* of 1916. It was adopted in 1923 by the Permanent Commission of the

League of Nations on Biological Standards, and as a result of the Frankfurt Conference in 1928 the international standard ouabain was prepared. Tincture of strophanthus is prepared in the United States from the seeds of *Strophanthus kombé* or *S. hispidus*, and in Great Britain from *S. kombé*. Now Jacobs and Hoffman (1926) have shown that the mixture of glycosides present in *S. kombé* consists of cymarin combined with different numbers of molecules of glucose, and such a mixture is quite different chemically from ouabain. Hence it is not surprising that when tinctures of *S. kombé* are tested in comparison with ouabain by different methods the results obtained depend on the method used. Burn and Trevan (1926) found, curiously enough, that results by the cat method and the frog method (*Rana temporaria*) were the same, but Edmunds, Lovell, and Braden (1929) found that tinctures of strophanthus appeared much more potent when tested on *Rana pipiens* by the 1-hour method. The results of the examination of a series of tinctures in comparison with ouabain have been described in a paper by Gaddum (1932) from which the figures in Table LXI are taken.

TABLE LXI. *Figures are mg. anhydrous ouabain to which 1 c.c. tincture is equivalent in activity*

Species	*Observer and method*	*Tincture A*	*Tincture B*	*Tincture C*	*Tincture D*	*Tincture E*
1. *Tests on frogs*						
R. pipiens	Smith *U.S.P.* method	6·5	7·9	6·9	7·2	6·5
R. pipiens	Smith 18-hr. method	5·0	4·5	4·0	5·0	6·1
R. temporaria	Gaddum 18-hr. method	3·5	3·4	3·0	3·8	5·0
R. temporaria	Trevan 18-hr. method	3·0	3·3	2·5	2·7	3·3
2. *Slow intravenous administration*						
Cat	Burn	3·1	3·4	2·3	2·8	3·1
Cat	Strangeways and Vartiainen	3·4	3·8	3·0	3·5	4·2
Dog	Tiffeneau	3·1	3·4	3·0	3·0	3·2
Guinea-pig	Knaffl-Lenz	2·1	1·95	2·6	2·5	2·1
Rabbit	Underhill	3·2	3·1	2·8	3·15	3·9

It is evident that a tincture when tested by the 1-hour method of the *United States Pharmacopoeia* may appear three times as strong in relation to ouabain as when tested by intravenous

injection into guinea-pigs. Hence the *British Pharmacopoeia* has adopted a Standard Tincture of Strophanthus, prepared from *S. kombé* seeds, so that unknown tinctures are compared not with ouabain, from which they differ qualitatively, but with a standard from which they differ only quantitatively. This Standard Tincture is equivalent in activity to a 0·42 per cent. solution of the international standard ouabain, or to a 0·33 per cent. solution of anhydrous ouabain. The international standard ouabain contains 20 per cent. of water of crystallization; it has the advantage of being stable in air, whereas anhydrous ouabain, or samples containing less water of crystallization, are said to be not so stable. The ouabain of the *United States Pharmacopoeia* contains about 12 per cent. of water of crystallization, and in view of the use of different ouabains the Permanent Commission on Biological Standardization has recommended that results be expressed in terms of anhydrous ouabain, whatever variety of ouabain is actually used.

Estimation by the frog method. For injection into *Rana temporaria* a dilution in 0·6 per cent. saline containing from 0·5 to 1·0 per cent. of tincture of strophanthus may be prepared; of this dilution 0·02 c.c. per gm. frog is injected into the lymph sac. The dose of ouabain is in the neighbourhood of 0·0005 mg. per gm., which may be administered by dissolving 2·5 mg. in 100 c.c. saline and injecting 0·02 c.c. per gm. It is best to dissolve the ouabain in 70 per cent. alcohol so as to make a 0·1 per cent. solution, and to dilute this further with saline as required.

Estimation by the cat method. To correspond with the concentration in which digitalis is usually administered to the cat tincture of strophanthus should be diluted 900 times. In Table LVIII it will be noticed that Macdonald diluted 500 times. Ouabain may be administered in a solution containing 0·5 mg. in 100 c.c.

Estimation by the guinea-pig method. Tincture of strophanthus should be diluted 250 times; ouabain may be administered in a solution containing 1·75 mg. in 100 c.c.

(ii) The Estimation of Strophanthin

A series of commercial samples of strophanthin *kombé* has been examined by Wokes (1928), who found that the potency

when tested by the cat method varied from 25 to 60 per cent. of that of the ouabain of the *United States Pharmacopoeia.* The average potency of all samples (excluding one unusually weak) was found to be 47 per cent. of this ouabain. Hence a useful guide to the dosage of strophanthin for the frog, cat, or guinea-pig methods is to take an amount of strophanthin equal to twice the quantity of ouabain already indicated.

Strophanthin B.P., 1932. In the *British Pharmacopoeia* an effort has been made to ensure that different samples of amorphous strophanthin have the same activity by requiring that all possess an activity which is 40 per cent. of that of anhydrous ouabain. Stronger samples are required to be diluted with lactose.

To test unknown samples of strophanthin a worker in Great Britain receives from the National Institute for Medical Research, London, not ouabain, but a solution of standard strophanthin; this was obtained from seeds of *Strophanthus kombé*, and is therefore qualitatively as similar as possible to any unknown sample of strophanthin obtained from *kombé* seeds (which alone are official). Unknown samples should differ only quantitatively from this standard, and the results of comparisons should be the same whatever the method used. The strophanthin used for preparing the standard was found to contain 59 per cent. of the activity of anhydrous ouabain by the frog method; the standard strophanthin as distributed is an alcoholic solution of which 1 c.c. contains activity equivalent to that of 5 mg. of the strophanthin defined in the *British Pharmacopoeia.*

III. The Estimation of Squill

Standard. No standard for squill has been adopted by the Permanent Commission on Biological Standardization. A standard may be prepared by mixing together equal volumes of a series of tinctures of squill, when the resulting mixture will represent a tincture of average potency. A composite tincture so prepared in the author's laboratory was found to be equivalent to a solution containing 0·013 per cent. of the glycoside Scillaren A when tested by the cat method.

When unknown samples of tincture of squill are compared with a composite tincture used as standard, the dilution and

dose of the tincture of squill should be the same as for tincture of digitalis. This applies to tincture of squill prepared as in the *British Pharmacopoeia*, 1932, of which 100 c.c. tincture contain the activity of 10 gm. squill.

IV. Estimation of Unofficial Preparations

Digitoxin. Digitoxin may be tested by any of the methods described. It is conveniently dissolved in a mixture of 6 parts of 90 per cent. alcohol, 3 parts of glycerine, and 1 part of distilled water. A 0·1 per cent. solution may be made, and this is further diluted with saline. Gage (1934) has estimated the potency of four samples in comparison with the international standard digitalis powder, and obtained results varying from 1,000 to 1,500 units per gm. These estimations were made by the frog method and also by the guinea-pig method.

Digitalin. Digitalinum Purum Germanicum is water-soluble and is commonly used in pharmaceutical practice for preparing hypodermic or intramuscular injections. Gage (1933) found that the activity of four samples varied from 55 to 250 units per gm., the activity of each sample being about the same both by frog and guinea-pig methods. The *British Pharmaceutical Codex* requires Digitalin to be standardized to contain 80 units in 1 gm.

V. Results by Different Methods

There is increasing evidence that when a sample of digitalis is compared with the international standard digitalis powder by different methods the result is not always the same. Thus a sample may be shown by the frog method to be 1·4 times as strong as the standard, while by the cat method it is only 1·2 times as strong. Again Wokes (1930) has shown that tincture of digitalis deteriorates at the rate of 3 per cent. per month when examined by the frog method, but that no deterioration is detected at all by the cat method. It is evident that the relative sensitivity to different digitalis glycosides differs in different species, and we are faced with the problem of discovering which method gives the right answer. There is no way of doing this save to make a careful comparison of different tinctures on large numbers of patients. Some attempts of this kind have been made, but the results are inconclusive.

Notwithstanding this difficulty, the biological standardization is valuable in preventing the use of adulterated leaf or leaf of very low activity for the treatment of heart disease.

REFERENCES

BERRY and DAVIS, 1935, *Quart. J. Pharm. Pharmacol.* **8**, 443.
BURN, 1930, *J. Pharmacol. Exp. Ther.* **39**, 221.
BURN and TREVAN, 1926, *Pharmaceut. J.* **117**, 493.
CHAPMAN and MORRELL, 1931, *Quart. J. Pharm. Pharmacol.* **4**, 195.
CHOPRA and CHOWAN, 1933, *Ind. J. Med. Res.* **20**, 1189.
DAVID and RAJAMANICKAM, 1934, *Quart. J. Pharm. Pharmacol.* **7**, 36.
EDMUNDS, LOVELL, and BRADEN, 1929, *J. Amer. Pharm. Ass.* **18**, 568.
EPSTEIN, 1933, *Quart. J. Pharm. Pharmacol.* **6**, 169.
GADDUM, 1932, ibid. **5**, 274.
GAGE, 1933, ibid. **7**, 654.
HANZLIK, 1929, *J. Pharmacol. Exp. Ther.* **35**, 363.
HATCHER and BRODY, 1910, *Amer. J. Pharm.* **82**, 360.
JACOBS and HOFFMAN, 1926, *J. Biol. Chem.* **67**, 609; **69**, 153.
KNAFFL-LENZ, 1926, *J. Pharmacol. Exp. Ther.* **29**, 407.
KROGH and HEMMINGSEN, 1926, *League of Nations Report on Insulin Standardisation*, C.H. 398.
DE LIND VAN WIJNGAARDEN, 1926, *Arch. exp. Path. Pharmak.* **113**, 40.
MACDONALD, 1934, *Quart. J. Pharm. Pharmacol.* **7**, 182.
MACDONALD and SCHLAPP, 1930, ibid. **3**, 450.
RAPSON and UNDERHILL, 1935, ibid. **8**, 409.
WOKES, 1928, ibid. **1**, 513.
—— 1930, ibid. **3**, 205.

CHAPTER XIV

ANTIPYRETICS AND ANALGESICS

Antipyretics. SUBSTANCES having an antipyretic action may be compared with one another by the methods described by Brownlee (1937, 1939). Cats may be used in which fever is induced by giving injections of a culture of *Bacillus coli communis*. Rats may be used in which fever is induced by the injection of a suspension of brewer's yeast.

Experiments on cats. Brownlee used a 96-hour peptone broth culture of two mixed strains of *Bacillus coli*, Houston 1 and Escherich, prepared to contain 1×10^{10} organisms per c.c. Though he does not say so, presumably the culture was killed by heating to 65° C. for 1 hour according to the description of Kiliani (1910) who introduced this method of producing fever. The normal temperature variation of cats kept in a room at 20° C. was found by Brownlee to be from 99·0° C. to 102·6° F. Cats weighing between 2 and 3 kg. were selected and kept without food, though given water, for 24 hours. The dose of culture was injected subcutaneously, giving 2·5 c.c. per kg. The rise of temperature was evident 1 hour after injection, and after 2 hours the temperature was sufficiently steady for the antipyretic test to be carried out. In one cat to which no antipyretic was given the initial rectal temperature was 100·6° F. Two hours after the culture was injected the temperature was 103·2°; at 6 and 7 hours, the temperature was 104·8° F. The pyrexia persisted for 24 hours. Experiments with antipyretics could be repeated in the same animals at weekly intervals. The antipyretic substance was administered orally from a graduated pipette. For example. in one experiment in which phenacetin was compared with aspirin, phenacetin was given in a dose of 33 mg. per kg. in 1 c.c. of 2 per cent. acacia suspension followed by an equal volume of tap water; or the drug was enclosed in a hard gelatine capsule which could be dipped in glycerine and placed at the back of the cat's tongue. When the cat's mouth was closed and the throat was stroked the capsule was swallowed. The capsule was followed by giving 5 c.c. tap water.

The test was carried out on two groups of cats. One group

received the substance used as a standard; the second grou received the unknown. The temperature was then recorded hourly for 3 hours after giving the drug.

Seven days later the second part of the test was performed in which the cats were crossed over so that the cats which received the standard in the first experiment received the unknown, and vice versa. The results of one experiment are given in Table LXII. The observations on cats A, B, and C given phenacetin were made at the same time as observations on cats D, E, and F given acetanilide and vice versa. The result for each cat was obtained by taking the differences between the temperatures at 1 hour, 2 hours, 3 hours and the temperature before the drug was given. These differences when added together give the figure in the last column but one of Table LXII.

TABLE LXII (*Brownlee, 1939*). *Testing antipyretics on Cats*

1. *Results when 33 mg. per kg. phenacetin given*

Cat	Rectal temperatures				Sum of differences	Mean
	Before	*1 hr. after*	*2 hrs. after*	*3 hrs. after*		
A	104·2	103·2	102·9	103·5	3·0	
B	104·1	102·6	102·6	103·3	3·8	
C	104·6	103·0	102·6	103·1	5·1	3·53
D	104·0	102·5	102·7	103·0	3·8	
E	104·8	103·5	103·7	103·8	3·4	
F	104·5	103·4	103·6	104·4	2·1	

2. *Results when 33 mg. per kg. acetanilide given*

Cat	Rectal temperatures				Sum of differences	Mean
	Before	*1 hr. after*	*2 hrs. after*	*3 hrs. after*		
A	104·5	100·7	101·5	103·0	8·3	
B	105·0	102·9	103·4	103·4	6·3	
C	106·0	102·2	102·2	103·6	10·0	8·73
D	104·2	102·2	99·4	100·2	10·8	
E	104·5	101·5	101·2	101·3	9·5	
F	104·5	102·2	101·8	102·0	7·5	

The two figures for the mean of the sum of differences indicate which substance is the more powerful antipyretic if both are given in the same dose. Thus the figures in Table LXII show that acetanilide is more potent than phenacetin, but do

not show how much more potent. To determine equivalent doses of the two substances, more experiments must be carried out.

Experiments with rats. Young male albino rats between 100 and 200 gm. were used. They were arranged in groups of eight animals and fed on a uniform diet until 24 hours before giving the drugs, when food was withdrawn. The room temperature was maintained throughout between 20° and 21° C. At 6 p.m. the rats were injected subcutaneously with a 15 per cent. suspension of dried brewers' yeast in 2 per cent. acacia in normal saline. The volume given was 1 c.c. per 100 gm. About 15 hours later the rectal temperature of each rat was taken and the drug administered by stomach tube suspended in 2 per cent. acacia.

The remainder of the test was carried out exactly as described for cats. The relative potency of different antipyretics was found to be the same for rats as for cats, as shown in Table LXIII.

TABLE LXIII. *Relative antipyretic potency* (*Brownlee*)

	Cats	*Rats*
Acetanilide	200	200
Amidopyrin	150	134
Phenacetin	100	100
Phenazone	100	100
Aspirin	60	60

Analgesic Action

Tests on human subjects. The problem of finding a widely accepted method for testing analgesic substances is still unsolved. The work of Wolff, Hardy, and Goodell (1940) on the pain threshold in human subjects has been followed by many attempts to determine the pain threshold by others. Thorp (1946) says that he has used this method but could not obtain the same results, although the apparatus used was exactly as described. He was unable to reproduce the time action curves for morphine sulphate with doses of 0·1–30 mg. for man. He points out that similar difficulties were met by Dodds, Lawson, Simpson, and Williams (1945) in testing diphenylethylamine compounds. The only explanation of these discrepancies

appears to be that some human subjects are specially sensitive to pain, and will therefore give results like those described by Wolff, Hardy, and Goodell, while others, who are perhaps the majority, cannot discriminate. It is certain that in making observations on man, the subject must not know what drug he is receiving.

Method of D'Amour and Smith. D'Amour and Smith (1941) used a method in which the rays from an electric light were focused on the tip of a rat's tail. The light was a 6–8-volt bulb, placed about 6 inches above the rat, the tail of which was lying in a groove. The light was turned on and a stop-watch was started to determine the time before the reaction, a sudden twitch of the tail, occurred. The authors used an intensity of illumination which produced a response in 5 seconds in the majority of animals. Intraperitoneal injections of morphine sulphate produced complete analgesia in a percentage of rats which depended on the dose. Thus,

6	mg. per kg.	produced	complete	analgesia in	none.
8	,,	,,	,,	,,	50 p.c.
10	,,	,,	,,	,,	66 p.c.
12	,,	,,	,,	,,	92 p.c.

When the end-point of complete analgesia is taken, this method appears to be quite reliable, but suffers from the disadvantage that it is not likely to detect minor degrees of analgesia, or to be of value for comparing different substances of the potency of aspirin. It is desirable to have a method in which doses of morphine not greater than 3 mg. per kg. are clearly effective.

In a later paper, Smith, D'Amour, and D'Amour (1943) attempted to distinguish degrees of analgesia short of complete analgesia by giving a score to each rat depending on the time taken to respond. The slowest response for untreated rats was 6·5 seconds; a response from 7 to 7·5 seconds was therefore given the score 1; a response from 8 to 8·5 seconds was given the score 3; a response from 9 to 9·5 seconds was given the score 6, a response not seen in 10 seconds was given the score 10. No longer exposure was used. Thus they were able to express the degree of analgesia for a group of 10 rats as the sum of the scores for each rat in the group. Even with this modification the authors found that little analgesia was conferred by analgesics such as aminopyrine or aspirin unless very large doses

were given. Otherwise the method as described by the authors appeared to be satisfactory.

Thorp's method. Thorp has described a method which is essentially the same as the method of D'Amour and Smith but in which the source of illumination has been modified from the Hardy, Wolff, and Goodell apparatus. The stimulus intensity is adjusted by the variable resistance controlling the lamp current, and the pain threshold represented by the power of the lamp in watts is determined by making several trials. New rats are said to be trained 'to discriminate between stimuli of different intensity before using them for the test. This is done by allowing several successive stimuli of high intensity to fall upon the tail until the rat removes its tail in a positive manner after each stimulus. The stimuli are repeated at a slightly lower intensity and the rats are then able to discriminate sufficiently accurately.'

The effect of an analgesic substance is determined as the percentage change in the initial pain threshold 30 minutes after the substance is given. In a typical experiment 3 groups of 3 rats are taken and each group is given a different dose. The next day the experiment is repeated, the same doses being given, but the groups are moved round. Finally, the experiment is completed on the third day, by which time every rat has received each dose, and the mean effect of each dose can then be calculated.

Hot-plate method. A method in which mice were put on a hot plate has been described by Woolfe and Macdonald (1944). A zinc plate was used which could be maintained at any temperature above 45° C. within ±0·5° C. On the plate was a hollow glass cylinder 15 cm. in diameter into which the mouse was dropped. When the plate was 55° C. all tested animals reacted in 30 seconds. The first sign of discomfort was that the mouse sat up on its hind legs, and when the pain was too great to be borne by the back paws, the mouse kicked its legs and danced about. The movement of the hind limbs was taken as the sign of acute discomfort. The standard time of exposure was 30 seconds. Groups of 10 mice were used, and if the dose of analgesic given subcutaneously was effective, it caused a proportion of mice in the group to tolerate the hot plate without reaction, the proportion depending on the dose. The duration of analgesia could be determined in relation to the dose at each

one of several temperatures of the hot plate. For example, when morphine hydrochloride was used, 5 mg. per kg. produced analgesia in 3 out of 10, while 7·5 mg. per kg. produced it in 8 out of 10 when the plate temperature was 55° C. The average duration of analgesia was 18 minutes for the dose of 5 mg. per kg., and was 37 minutes for the dose of 7·5 mg. per kg. However, with this method, too, aminopyrine and antipyrin were effective in very large doses only.

Stimulation by heat. A method has been described by Davies, Raventos, and Walpole (1946) in which a heat stimulus is applied to the rat's tail. The rat is placed in a cylindrical holder and its tail lies along a channel made by placing two strips of asbestos board (uralite) on top of a sheet of asbestos board. A hole is made in the board not more than 1½ inches from the tip of the tail, and on the other side of the hole is a small coil of resistance wire connected through a key to a 6-volt supply. The wire is of such a gauge and length that with the circuit closed it is raised to a bright-red heat. The reaction time is determined by a stop-watch from the moment of closure of the key to the withdrawal of the tail. The heating coil is adjusted so that the majority of animals react in 5 seconds. The effect of analgesics is measured by their effect in prolonging the withdrawal time.

Method of Reinhard and de Beer. The methods of Thorp and also of Davies, Raventos, and Walpole have both been carefully tried in the author's laboratory without much success. The rat's tail does not seem to be a sufficiently sensitive organ of sensation; the thick tail of a full-grown rat is developed for other purposes, and it is surprising that the authors of these methods have not adopted the use of young rats which have a more delicate skin.

We have obtained better results by using a method described by J. F. Reinhard and E. J. de Beer (Wellcome Research Laboratories, Tuckahoe, N.Y.), based on the electrical stimulation of the mouse tail, about which we were privately informed by Dr. de Beer. The method has not been published because the authors were not satisfied with their own results, but with their permission this brief account is now being given. The mouse is placed in a holder shown in Fig. 51, and its tail is cleaned with ether. Some electrode jelly is rubbed into the tail, taking care to leave no surplus of jelly along the tail through

which the current can short. The electrodes used in this laboratory consisted of pieces of clockspring 1 cm. wide, with weights on the ends to keep them pressed on the tail. The electrodes are dropped across the tail, the space between them being 1·5 cm.

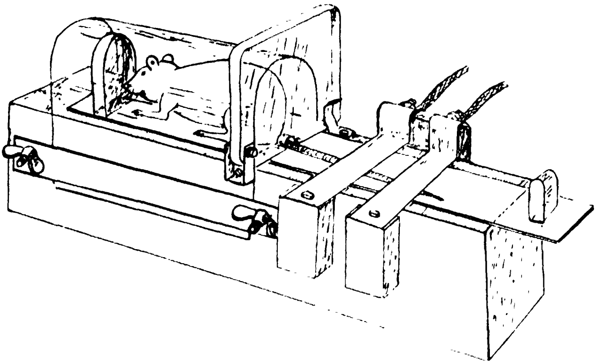


FIG. 51. Apparatus of Reinhard and de Beer for applying stimuli to mouse tail in test for analgesia.

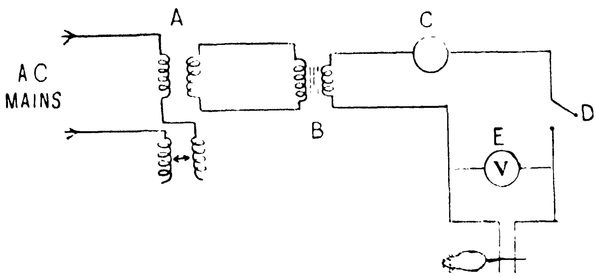


FIG. 52. *A* variac; *B* small 3/1 transformer; *C* contact make and break, revolves 1 per sec. and gives shock lasting 1/26 sec.; *D* key; *E* voltmeter (reading 6 volts).

The key is then closed and the number of shocks needed to make the mouse squeak is counted. Each shock is indicated by a backward flick on a voltmeter.

The circuit used is indicated in Fig. 52. Leads from A.C. mains go to a variac and then to a small 3/1 transformer. There is a key in the circuit from the transformer to the electrodes which are applied to the tail and there is also a rotating contact which rotates once a second. and completes the circuit each time for a duration of 1/26 second. A voltmeter is in circuit

across the electrodes and the variac is adjusted to give a current of 6 volts across the electrodes. As a rule mice squeak after 3 shocks; the key is at once opened. Mice which did not squeak within 5 shocks were not used.

The drug was administered orally in a volume of 0·2 to 0·5 c.c. to mice which had been kept without food overnight. The maximum effect was 15 minutes after administration.

The effect shows itself as an increase in the number of shocks which must be applied before the mouse squeaks. Reinhard and de Beer expressed the degree of analgesia in the following way. A control group of 10 mice squeaked after 2·6 shocks (mean number). When they received 1·25 mg. per kg. morphine, another group of 10 mice squeaked after 5·9 shocks. In one of the mice which received the morphine the number of shocks was 4. The proportion of this score due to the drug was

$$4 \times \frac{5{\cdot}9 - 2{\cdot}6}{5{\cdot}9} = 2{\cdot}24.$$

2·24 is the corrected score. The authors found a linear relation between log dose and log of the mean corrected score.

On trying this method in the Oxford laboratory we found it possible to obtain clear evidence of the analgesic effect of phenazone and similar weak analgesics such as papaverine. The work was carried out by J. R. Vane, who considered that one drawback lay in the variation in the resistance of the different mouse tails which he thought was responsible for the wide scatter of the scores after dosing the mice with an analgesic. To lessen this he proposed to insert a resistance in series with the electrodes of such a size that the variation in the resistance of the tails would be negligible, and then to increase the voltage so that the current passing through the tail would be about the same as before.

The writer's opinion is that perhaps it may be better to use the method simply to obtain evidence of the presence of analgesia and not to measure the analgesia in each mouse. If any score greater than five or six is taken as evidence of analgesia it should then be possible to relate the dose of analgesic to the percentage of mice in which analgesia is found present, and thus to find the dose causing analgesia in 50 per cent. of mice. In this way the potency of different analgesics could be compared.

REFERENCES

BROWNLEE, 1937, *Quart. J. Pharm. Pharmacol.* **10**, 609.
---- 1939, ibid. **12**, 45.
D'AMOUR and SMITH, 1941, *J. Pharmacol.* **72**, 74.
DAVIES, RAVENTOS, and WALPOLE, 1946, *Brit. J. Pharmacol.* **1**, 255.
DODDS, LAWSON, SIMPSON, and WILLIAMS, 1945, *J. Physiol.* **104**, 47.
KILIANI, 1910, *Arch. internat. pharmacodyn.* **20**, 333.
SMITH, D'AMOUR, and D'AMOUR, 1943, *J. Pharmacol.* **77**, 184.
THORP, 1946, *Brit. J. Pharmacol.* **1**, 113.
WOLFF, HARDY, and GOODELL, 1940, *J. Clin. Invest.* **19**, 659.
WOOLFE and MACDONALD, 1944, *J. Pharmacol.* **80**, 300.

CHAPTER XV

ATROPINE SUBSTITUTES

Mydriasis. Suppression of Secretion

Test on the mouse pupil. The best method for making quantitative comparisons of different mydriatic substances is that of Pulewka (1932). The only apparatus required is a binocular microscope magnifying about ten times and provided with a scale in the eyepiece with which to measure the diameter of the pupil of the mouse. A light which shines into the eye of the mouse is attached to the microscope. Groups of mice are taken preferably from 15 to 20 gm. in weight; larger mice are less sensitive, but they may be used provided that all the mice in any experiment are of similar weight. They are not kept for any time without food. Pulewka recommended using mice of one sex, but this is not necessary. The mice are injected intraperitoneally with 0·2 c.c. of the solution and the effect on the pupil is determined. The action of a dose of atropine reaches its maximum in 15 minutes and begins to decline after 25 minutes; for this reason readings are taken within 15–20 minutes after the injection.

Some results given in a paper by Ing, Dawes, and Wajda (1945) were as follows. The diameter of the pupil in uninjected mice was found to vary between 1·5 and 2·5 divisions of the scale in the microscope eyepiece (7·7 divisions = 1 mm.); this diameter was unaffected by excitement. After the injection of 2 μg. atropine sulphate the average diameter was usually about 4 divisions, while after 8 μg. the diameter was about 14 divisions, and within the range 2–8 μg. the mean diameter of the pupil was directly proportional to the logarithm of the dose, as is shown in Fig. 53. The points were obtained by combining 10 experiments, each point representing the mean of results in 100 mice.

The effect of a given dose of atropine sulphate varied in different groups of mice, and even within the same group at different times. For this reason, in comparing a substance of unknown potency with atropine, the effect of two or three different doses of atropine sulphate in different groups of mice

was always determined simultaneously with that of a range of doses of the substance under comparison. By plotting the results obtained with atropine as in Fig. 53, the effect of a dose of the unknown could be equated to that of a dose of atropine sulphate.

In practice it was found convenient to use 5 mice in a group,

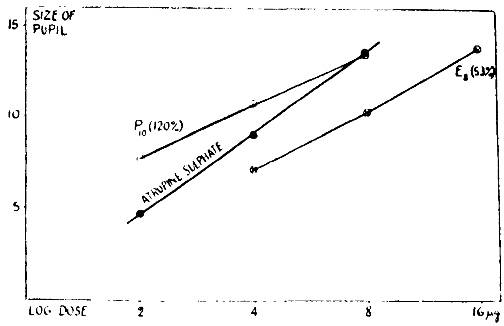


FIG. 53. Log dose—action curves for atropine and synthetic mydriatics. Ordinates: diameter of mouse pupil in divisions of microscope scale. Abscissae: logarithm of dose in micrograms. (Ing, Dawes, and Wajda, *J. Pharmacol.* 1945, **85**, 88.)

and to inject 3 groups. The diameters of the pupils were then determined. The experiment was continued by injecting 3 more groups and so on. The details of part of an experiment in which l-hyoscine hydrobromide was compared with atropine sulphate are given in Table LXIV. The part illustrated occupied one

TABLE LXIV

Substance	*Dose per mouse*	*Time of* *Injection*	*Time of* *Examination*	*Pupil diameters*	*Mean*
	mg.				
Atropine sulphate	0·002	11·18	11·33	2, 3, 2, 7, 5	3·8
	0·004	11·21	11·36	4, 6, 9, 2, 11	6·4
	0·008	11·24	11·39	15, 15, 15, 4, 17	13·2
Hyoscine hydrobromide	0·00025	11·51	12·06	4, 2, 2, 3, 5	3·2
	0·0005	11·55	12·10	7, 4, 6, 7, 6	6·0
	0·001	11·58	12·13	12, 12, 14, 12, 8	11·6

worker about 1 hour, excluding the time taken to prepare the solutions; the complete experiment involved repeating the same

doses twice on fresh groups of mice so that the effect of each dose of each substance was observed in 15 mice. Mice were not used more often than twice in 1 week, and after 3–4 weeks' use they were discarded because they grew bigger and less sensitive. The method can be used for comparing the duration of the mydriasis produced by different compounds. Observations on the size of the pupil are made at varying intervals after the injection.

Test on salivary flow. A method for the quantitative comparison of atropine substitutes on the salivary secretion of the cat has been published by Bülbring and Dawes (1945). This method depends on the observation that the inhibitory action of belladonna alkaloids is greater when the salivary flow is rapid, and the substance used to stimulate the flow is therefore infused intravenously to give as fast a flow as possible consistent with a constant rate of secretion. Under these conditions the duration of action of atropine is far shorter, and many observations may therefore be made on the same animal.

Method. Cats are used, anaesthetized with pentobarbitone or chloralose. Both anaesthetics give equally satisfactory results. A cannula is tied into Wharton's duct and attached to a bottle containing tap-water. The tap-water, displaced by saliva, passes out of the bottle through the tube which actuates a drop timer (Gaddum and Kwiatkowski, 1938). As each drop forms at the mouth of the tube it makes a contact which brings an electromagnet into action to record a vertical line on a drum. When the drop falls the contact is broken, and the vertical line is interrupted; the height of the line is thus proportional to the interval between one drop and the next. Records such as those seen in Fig. 54 are obtained. A steady flow of saliva is produced in the cat by an intravenous infusion at a uniform rate of a 0·004 per cent. solution of carbaminoylcholine in Ringer-Locke. Since this lowers the blood-pressure, adrenaline hydrochloride 0·002 per cent. is also included in the infusion. Stavraky (1942) has shown that adrenaline increases the salivary flow produced by muscarine-like substances. The injection of atropine substitutes not only slows the secretion but also causes a rise of blood-pressure, and a comparison of different substances can be made by both these effects.

Some form of constant infusion device must be employed

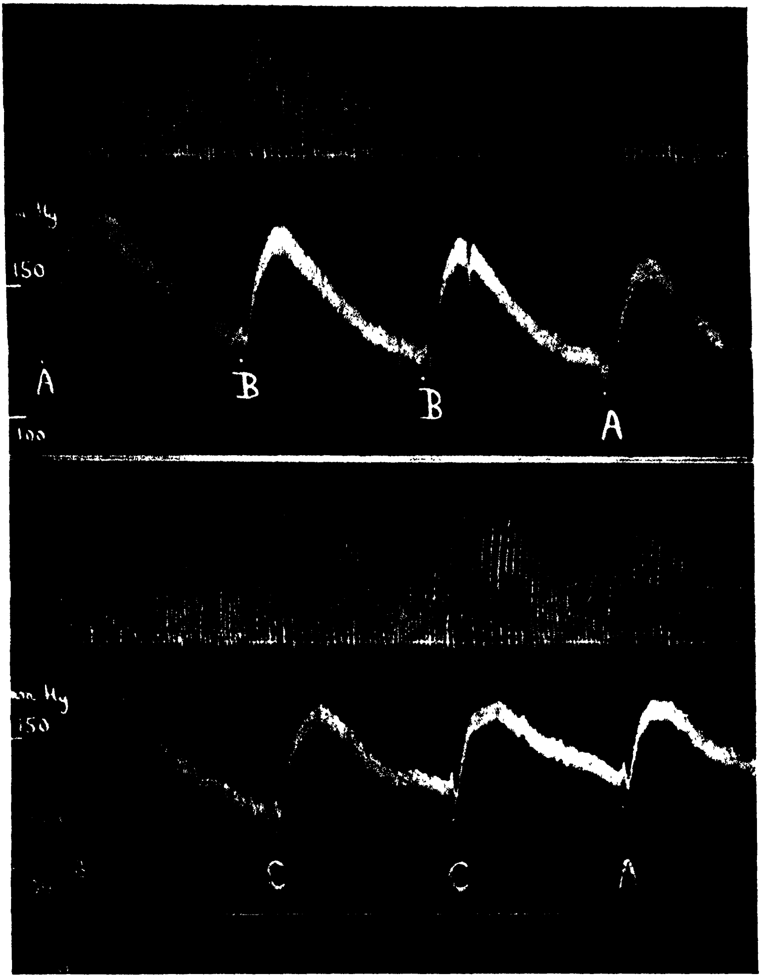


FIG. 54. At *A* intravenous injection of 2 μg. atropine sulphate, at *B* 1 μg. l-hyoscyamine sulphate, at *C* 0·7 μg. l-hyoscyamine methiodide. (Bülbring and Dawes, *J. Pharmacol.* 1945, **84**, 182.)

which is accurate and which can work for several hours without having to be stopped for refilling. Bülbring and Dawes used a pump which delivered liquid paraffin into a flask from which the solution of carbaminoylcholine and adrenaline was displaced.

The rate of infusion of carbaminoylcholine is usually about 0·5 c.c. per minute, but it requires adjustment for each animal

until optimal salivary flow is obtained. This is such that the intravenous injection of atropine sulphate in a dose of about 1–5 μg. at intervals of 5 minutes no longer causes a gradual decrease in the rate of secretion over a long period. This optimal condition is only reached after some trouble; in practice it is best to run the infusion with intermittent injections of atropine for half an hour, and then to stop it for 45 minutes. At the end of this time results are more constant, the resting blood-pressure steadier, and small doses of atropine can be injected every 3–5 minutes, causing a transient rise of blood-pressure and decrease of salivary secretion, without cumulative action. Fig. 54 shows a comparison of atropine sulphate with l-hyoscyamine sulphate and also with l-hyoscyamine methiodide.

REFERENCES

Bülbring and Dawes, 1945, *J. Pharmacol.* **84**, 177.
Gaddum and Kwiatkowski, 1938, *J. Physiol.* **94**, 87.
Ing, Dawes, and Wajda, 1945, *J. Pharmacol.* **85**, 85.
Pulewka, 1932, *Arch. f. Exper. Path u. Pharmakol.* **168**, 307.
Stavraky, 1942, *Rev. Canad. Biol.* **1**, 64.

CHAPTER XVI

LOCAL ANAESTHETICS

Surface anaesthesia. FOR many years there was no satisfactory method for measuring the potency of local anaesthetics, and much time was spent in making approximate determinations of the minimal anaesthetic concentration. Some progress has, however, been made in the last few years.

Chance and Lobstein (1944) have described a method of determining the median effective dose of a local anaesthetic when applied to the cornea of a guinea-pig. They have found that the guinea-pig is a better animal for work on the cornea than the rabbit. The general idea of the method is to apply a known concentration of a local anaesthetic to both eyes of a guinea-pig, and to test each eye five times during the next 5 minutes. The degree of anaesthesia is expressed as the number of times out of ten the blink reflex remains absent. The same solution of anaesthetic can be tested on several other guinea-pigs, and thus the mean percentage response (the mean percentage failure to blink) can be determined. By determining the mean percentage response for other concentrations of the anaesthetic, and plotting the logarithm of the concentration against the probit of the percentage response, the concentration causing a 50 per cent. response can be determined. This is the median effective dose.

In applying the solution to the eye of the guinea-pig, Chance and Lobstein hold the guinea-pig by a clamp which exerts light pressure on the flanks behind the ribs. The solution is dropped on to the cornea so that the space between the eyelids contains a clearly visible film of solution. If it is expelled by a blink the application is repeated, and if necessary a third application is made. The first stimulation is made 45 seconds after the instillation. A horse-hair mounted on a glass rod is used which is pressed against the centre of the cornea so that the hair is bent to the same extent at each application.

Results in which cocaine, procaine, and amethocaine (butethanol) were compared by this method are given in Table LXV.

Intradermal anaesthesia. A test in some respects similar to

TABLE LXV. *The median effective concentrations, E, in mg. per c.c. aqueous solution, using the hydrochlorides*

	Amethocaine	*Cocaine*	*Procaine*
E	0·56	2·63	26·0
True fiducial limits for $P = 0{\cdot}95$.	0·30–0·67	1·73–3·51	12·6–29·0

the above has been devised by Bülbring and Wajda (1945) in which the anaesthetic solution is injected intracutaneously in guinea-pigs. The area of injection is outlined in ink. The time of injection is noted, and a test for anaesthesia is then applied to the injected area at regular intervals of 5 minutes for 30 minutes. The test consists in applying 6 pricks to the area and noting how many of these pricks fail to elicit a contraction of the surrounding skin. A quantitative expression for the intensity of anaesthesia is thus obtained.

Details of test. Fully grown guinea-pigs are chosen. On the day preceding the experiment the hair on the back is clipped and two areas of 4–5 cm. diameter are shaved. This produces a certain amount of irritation which disappears overnight. The sensitivity of the skin is greatest in the midline and slightly more in the front area than in the back area. For this reason each concentration of a local anaesthetic must be tested in both areas. Six tests using 3 guinea-pigs can be performed simultaneously. The dose is always injected intracutaneously in 0·25 c.c. saline. Three guinea-pigs receive one dose in the front area and another dose in the back area; the size of the wheal is marked with ink. The reaction to pin-prick is tested 5 minutes after the injection in the following way. After observing the animal's normal reaction to a prick applied outside the wheal, 6 pricks are applied inside the wheal and the number of pricks is counted to which the guinea-pig fails to react. The pricks are applied at intervals of about 3–5 seconds. The test of 6 pricks is applied every 5 minutes for 30 minutes. The number of times the prick fails to elicit a response during the 30-minute period is added up and the sum, out of a possible 36, gives an indication of the degree of anaesthesia. Having completed the test on 3 guinea-pigs, the same solutions are injected into 3 other guinea-pigs, but the solution which was used for the front

is now used for the back area and vice versa. An experiment shown in Table LXVI determining the relation between different concentrations of nupercaine and the degree of local anaesthesia may serve as an example. Table LXVI shows that an average response from 6 tests was obtained for each concentration. After injecting 0·0125 per cent. nupercaine the guinea-pigs failed to respond to pin-prick only 10 times out of 36, whereas 0·025 per cent. produced 17 failures, 0·05 per cent. produced 25 failures, and 0·1 per cent. produced almost full anaesthesia for 30 minutes.

TABLE LXVI. *Number of pricks (out of six) failing to elicit a response after intracutaneous injection of nupercaine in guinea-pigs*

	0·0125 per cent.						*0·025 per cent.*					
	Back			*Front*			*Front*			*Back*		
Time	*Pig 1*	*Pig 2*	*Pig 3*	*Pig 4*	*Pig 5*	*Pig 6*	*Pig 1*	*Pig 2*	*Pig 3*	*Pig 4*	*Pig 5*	*Pig 6*
min.												
5	5	6	6	5	3	4	6	6	6	6	6	6
10	3	4	2	2	3	4	3	6	5	3	5	6
15	3	2	2	0	0	3	3	2	5	2	3	4
20	1	0	1	..	..	0	3	1	4	0	3	4
25	0	..	0	..	..	..	1	0	0	..	3	3
30	..	..	..	..	..	..	0	..	..	..	0	0
	Mean sum for 30 minutes 9·8 ± 1·08						Mean sum for 30 minutes 17·5 ± 1·76					

	0·05 per cent.						*0·1 per cent.*					
	Front			*Back*			*Back*			*Front*		
Time	*Pig 7*	*Pig 8*	*Pig 9*	*Pig 10*	*Pig 11*	*Pig 12*	*Pig 7*	*Pig 8*	*Pig 9*	*Pig 10*	*Pig 11*	*Pig 12*
min.												
5	6	6	6	6	6	5	6	6	6	6	6	6
10	4	6	6	6	5	4	6	6	6	6	6	6
15	4	5	5	6	5	4	6	6	6	6	6	6
20	4	4	3	6	5	4	6	6	6	6	6	6
25	1	4	2	6	4	3	6	6	6	6	5	6
30	1	1	1	2	3	3	5	6	6	6	3	6
	Mean sum for 30 minutes 25·3 ± 1·86						Mean sum for 30 minutes 35·2 ± 0·66					

When these results for nupercaine are plotted against the log dose a straight line is obtained, this line being approximately parallel to similar lines for β-eucaine and procaine (see Fig. 55). The line for cocaine is, however, steeper, which makes it difficult to state the relative strength of the other anaesthetics to cocaine. In order to compare with cocaine, tests are carried out in the following way. A determination of the effect

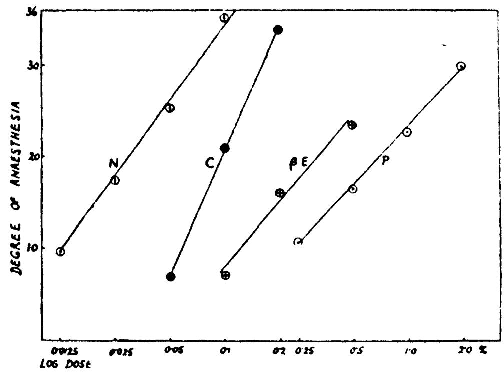


FIG. 55. Intracutaneous wheal in guinea-pigs. The graph shows the relation between the concentration of the local anaesthetic (abscissae) and the number of times there was no response to a prick; total 36 pricks in 30 minutes (ordinates). For nupercaine (*N*) each point represents the mean of 6 observations, for cocaine (*C*) of 33 to 90 observations, for β-eucaine (β-*E*) of 9–15 observations, for procaine (*P*) of 12–21 observations. (Bülbring and Wajda, *J. Pharmacol.* 1945, **85**, 81, Fig. 2.)

of two or three different concentrations of cocaine was always made on the same day as that on which the compound was tested. Two doses were chosen of the latter which were likely to have effects intermediate between the effects of cocaine. By plotting the results obtained with cocaine against the logarithm of the concentration used the relative potency of each concentration of the tested compound was determined graphically. Thus several figures were obtained in one experiment, and by taking the average an approximate value could be calculated. An example of such a test is given in Table LXVII.

Plexus anaesthesia in frogs. Bülbring and Wajda (1945) have developed this method originally introduced by Sollmann (1918)

TABLE LXVII. *Potency of β-eucaine in terms of cocaine* (= 100)

Intracutaneous injection in guinea-pigs

Concentration	*Effect*	*Per cent. potency*
0·05% cocaine	3·3 (3)±1·3	..
0·1% ,,	16·2 (6)±1·3	..
0·2% ,,	33·3 (3)±1·4	..
0·1% β-eucaine	4·0 (3)±1·4	53·8 } 51·3
0·2% ,,	16·8 (6)±1·5	48·9 }

Figures in brackets are numbers of animals.

in the following way. *Rana temporaria* are used which are kept in store at 4° C. No attention is paid to weight or sex. The frog is decapitated and the upper part of the spinal cord is destroyed down to the level of the third vertebra. A transverse incision is made in the abdominal wall just below the sternum. The viscera are removed through this opening, carefully exposing the lumbar plexus without damaging it. The frog is pinned to a vertical board. The solution of the local anaesthetic dissolved in 0·7 per cent. saline is put into the pocket formed by the lower abdomen. The amount of solution used is irrelevant as long as the plexus is submerged. A record is made of the time taken to abolish the reflex contraction to a sensory stimulus. No account is taken of the reflex time itself. The sensory stimulus is the immersion of both the frog's feet once every minute into dilute HCl for not longer than 10 seconds, after which the feet are rinsed in water. The first solution used as stimulus is 0·05 N HCl, and when the frog fails to react a stronger solution, 0·1 N HCl, is tested, and after that 0·2 N HCl is tested. At this point stronger acid also fails to elicit withdrawal of the foot. The end-point is determined in this way to avoid damaging the sensory nerve-endings by repeated application of stronger acid than necessary. Three or four frogs can be tested simultaneously and for each the observation is made of the time taken by a given concentration of local anaesthetic to abolish the reaction to 0·2 N HCl. Table LXVIII illustrates an experiment carried out with three different concentrations of cocaine. In this method also the relation between log concentration and time to anaesthesia is linear as shown in Fig. 56, but here also the slope of the cocaine line differs from that of the other anaesthetics.

TABLE LXVIII. *Time taken for the development of plexus anaesthesia in frogs*

Concentration of cocaine	*Frog*	*Time for anaesthesia to dilute* HCl *0·05* N	*0·1* N	*0·2* N	*Mean time for anaesthesia to 0·2* N HCl
Per cent.		*Min.*	*Min.*	*Min.*	
0·05	1	3	9	17	20·25 ± 4·3 min.
	2	7	18	23	
	3	4	28	28	
	4	2	9	13	
0·1	5	4	6	8	9·7 ± 1·0 min.
	6	4	8	11	
	7	3	4	12	
	8	5	7	8	
0·2	9	2	3	5	4·5 ± 1·3 min.
	10	1	7	8	
	11	0	0	2	
	12	1	2	3	

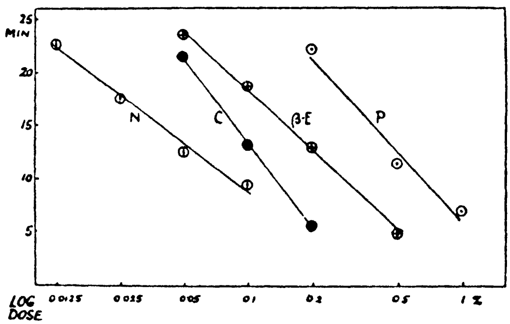


FIG. 56. Plexus anaesthesia in frogs. The graph shows the relation between the concentration of the local anaesthetic (abscissae) and the time of onset of anaesthesia in minutes (ordinates). N = nupercaine, C = cocaine, β-E = β-eucaine, P = procaine. Each point represents the mean of 6–12 observations. (Bülbring and Wajda, *J. Pharmacol.* 1945, **85**, 81, Fig. 1.)

Hence, in comparing an unknown substance with cocaine, it is best to obtain three points on the cocaine line and then to determine the effects of two concentrations of the unknown, these

being intermediate between the cocaine effects. An example is given in Table LXIX. The potency of procaine, β-eucaine, and

TABLE LXIX. *Potency of β-eucaine in terms of cocaine* (= 100)

Plexus anaesthesia in frogs

Concentration	*Effect*	*Potency per cent.*
0·05% cocaine	20·2 (4) ± 3·3	. .
0·1% ,,	9·5 (4) ± 1·3	. .
0·2% ,,	5·0 (6) ± 0·9	. .
0·1%-eucaine	17·2 (4) ± 1·4	54·9 } 51·3
0·2% ,,	11·5 (4) ± 2·0	47·7 }

nupercaine in relation to that of cocaine by both methods is given in Table LXX.

TABLE LXX. *Potency in terms of cocaine* (= 100)

	Intracutaneous wheal (guinea-pigs)					*Plexus anaesthesia (frogs)*				
Substance	*Lowest*	*Highest*	*Mean*	*Number of animals used for substance tested*	*Number of animals used for cocaine*	*Lowest*	*Highest*	*Mean*	*Number of animals used for substance tested*	*Number of animals used for cocaine*
			Per cent.					*Per cent.*		
Procaine .	8·5	17·4	13·5	18	30	40·5	42·4	41·5	12	17
β-eucaine .	48·9	53·8	51·3	9	12	47·7	54·9	51·3	8	14
Nupercaine.	225	315	282·5	12	12	140	224	169·5	14	14

REFERENCES

BÜLBRING and WAJDA, 1945, *J. Pharmacol.* **85**, 78.
CHANCE and LOBSTEIN, 1944, ibid. **82**, 203.
SOLLMANN, 1918, ibid. **11**, 1.

CHAPTER XVII

GASTRIC SECRETION

THE comparison of substances which affect gastric secretion has been recently described by Howat and Schofield (1948). They have shown how to compare substances which modify the secretion of hydrochloric acid produced by histamine in cats.

The cats should have been kept without food overnight. They are anaesthetized with ether followed by chloralose (80 mg. per kg.) given intravenously. The procedure proposed by Howat and Schofield has been modified in the following way. A small incision is made in the abdominal midline through which the pylorus is drawn. Two string ligatures are passed under the pylorus and the one farthest from the stomach is tied. An incision is made immediately above this tie and through it a rubber tube on the end of a glass tube is passed into the stomach; the second ligature is tied round the pylorus and glass tube. The rubber tube has several holes cut in the side so that gastric juice can readily enter it. Care is taken that the ligatures do not include any important blood-vessels. A second glass tube is passed into the cardiac end of the stomach through an incision in the oesophagus. The external end of this tube is attached to a funnel and the stomach is washed out with warm water.

If it is desired to exclude vagal stimulation of juice, atropine sulphate, 1 mg. per kg., is injected intravenously. The cat is then left for 1 hour, after withdrawing the tube from the oesophagus and tying this off.

The stomach is then washed, by way of the tube through the pylorus, with N/200 HCl, and 25 c.c. is left in. At the end of 15 minutes this is removed and titrated for free and total acid with N/20 NaOH. Another 25 c.c. N/200 HCl is replaced in the stomach and again removed after 15 minutes, and so on.

A flow of acid gastric juice is stimulated by intravenous infusion of histamine acid phosphate in 0·9 per cent. NaCl at a constant rate for 45 minutes. The rate of infusion recommended by Howat and Schofield is 0·0045 mg. per kg. per minute. This corresponds to 1·5 μg. histamine per kg. per minute, since the

base is one-third of the weight of the salt. When the secretion of acid has returned to the basal level, the same infusion of histamine for 45 minutes is repeated, and a second response of the secreting glands is obtained in the same way. An example is shown in Fig. 57 which is taken from an experiment carried out in the author's laboratory by N. K. Dutta; in this experi-

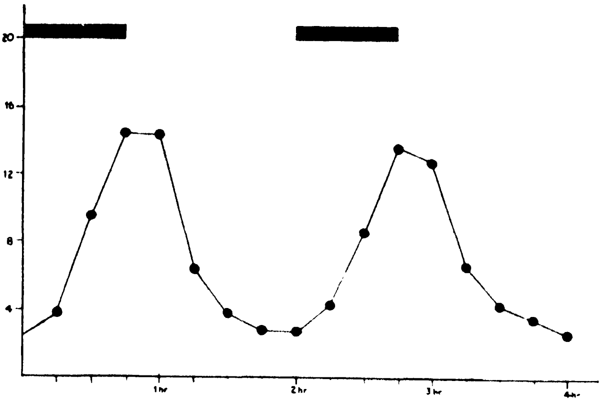


FIG. 57. Secretion of gastric HCl by cat in response to histamine infusion during periods indicated by black bars. Method of Howat and Schofield. (Experiment performed by N. K. Dutta) ordinates: c.c. free HCl N/20. Abscissae: time.

ment the rate of histamine infusion was 4·5 μg. histamine base per kg. per minute.

In Howat and Schofield's work the difference between the first and second responses is expressed as a percentage of the first, and in a series of experiments the mean of these percentage differences is calculated. The substance to be tested is injected intravenously just before the second histamine infusion begins. In a control series of twenty-five experiments with no intervening injection the mean percentage difference was found by Howat and Schofield to be −10·8 per cent.

In using the method a series of control experiments should be carried out, e.g. in ten cats. Experiments should then be made, injecting the substance in a given dose between the two histamine infusions, in another series of cats which should probably include not less than five cats, depending on the clearness of the effect of the substance tested. As is seen in Table LXXI, the effect of the second histamine infusion relative to

that of the first varies from cat to cat, and therefore results from one or two cats only may be misleading. The results given in the table were obtained by N. K. Dutta.

TABLE LXXI

Exp.	*Total amount of free acid secreted during 45 min. infusion of histamine.* *c.c.* N/20 HCl *1st response*	*2nd response*
1	15·2	26·1
2	8 3	14·3
3	22·7	16·7
4	15·4	16·4
5	11·5	14·3
6	14·6	27·8
7	26·6	19·1
8	15·3	14·8
9	25·4	23·9
10	17·2	15·0
Mean	17·2	18·8

REFERENCE

HOWAT and SCHOFIELD, 1948, *J. Physiol.* **107**, 30 P.

CHAPTER XVIII

SECRETIN AND PANCREOZYMIN

Standard preparation. No international standard for secretin or pancreozymin is available at the present time, but a standard preparation can be made in the form of an acetone-dried powder (Burn and Holton, 1948). The small intestine, preferably of a dog, is excised, washed quickly under the tap to remove blood, and a portion is taken. The first 30 cm. provides the most active powder. It is slit longitudinally and the mucosa washed with cold water. The mucosa is scraped off with a blunt scalpel, minced in a machine, and weighed on paper. It is transferred to a beaker and put in a freezing chamber. The procedure thus far should be completed within 10 minutes. When frozen solid, the mince is ground in a mortar surrounded by ice and salt. Ice-cold acetone is poured on, mixed, and the mixture is filtered on a Buchner funnel, using a filter-pump. The extraction is performed three times, after which the mucosa should appear finely divided. The material is spread in a dish and left in an evacuated desiccator over $CaCl_2$. After 23 hours the dry material is ground in a mortar and passed through a No. 100 sieve. The fine powder is now extracted with acetone in a Soxhlet apparatus for 4 hours and kept *in vacuo* over P_2O_5 for 24 hours. It is then distributed into small glass tubes with narrow necks and these are left *in vacuo* over P_2O_5 for 2 days. The desiccator is filled with dry air and the tubes are taken out and sealed at once.

An extract is prepared before each experiment by mixing 1 gm. of the powder with a small portion of 25 ml. 0·15 N HCl in an evaporating dish to make a smooth paste. The rest of the 25 ml. is added. The dish is put over a low flame and the contents brought to boiling-point in about 5 minutes. The mixture is stirred and allowed to boil for exactly 1 minute. The flame is removed, a piece of red litmus is added, and normal NaOH (usually 3·5 ml.) is run in with stirring until the protein mass coagulates and the litmus paper turns purple. The time taken for neutralization is 30 seconds. The extract is filtered through a folded paper into a 25 ml. measuring cylinder. The protein

mass and filter paper are washed with a few ml. of distilled water and the washings filtered into the measuring cylinder. The volume of filtrate is made up to 25 ml. so that 1 ml. corresponds to 40 mg. acetone dried powder. This extract loses activity in 24 hours.

Preparation of the cat. The cat, which has not been fed for 18 hours, is anaesthetized with chloralose (after induction of anaesthesia with ethyl chloride and ether) and the abdomen is opened by a midline incision of about 10 cm. from the xiphisternum. Hooks and weights are used to retract the abdominal wall. The pylorus and first duodenal loop are brought outside and turned towards the animal's left side. The bile-duct is identified by gently squeezing the gall-bladder. The pancreatic duct usually enters the duodenum about 2 mm. below the entry of the bile-duct. Starting about 2 cm. below the bile-duct, the pancreas is separated from the duodenum by blunt dissection, the blood-vessels being tied and divided where necessary. When the pancreatic duct is reached two silk ligatures are put round it and a length of 5 mm. is dissected. A dose of secretin is then injected intravenously and one of the ligatures is then tied as near the duodenum as possible. The duct is then cut and a flow of juice is usually visible. A small cannula is tied into the duct and attached to a rubber tube and L-piece whose position is adjusted so that the flow of juice continues. The cannula and L-piece are clamped in position and the abdomen is closed after removing the hooks and weights.

Estimation of secretin. The relation between the dose injected and the number of drops of pancreatic juice secreted has been found to be linear, but the slope of the line differs in different cats, and the amount of one preparation which is just sufficient to produce some secretion differs for different cats; it may be three times as great in one cat as in another. These points are illustrated in Fig. 58.

In making a comparison between two preparations, the dose-response relation should first be determined for one preparation by finding the amount of secretion in drops produced by three or more different doses of it. The line relating the response to the dose is then drawn in a graph. A dose of the second preparation is then injected, and, having observed the response to this, it can at once be seen from the graph what dose of the first

preparation would have produced the same response as the dose of the second. Further observations of the effect of the second preparation are then made to check the first estimate of the relative potency of the two preparations.

Estimation of pancreozymin. Pancreozymin is the hormone obtained from intestinal mucous membrane by Harper and Raper (1943), which stimulates an output of enzymes in pan-

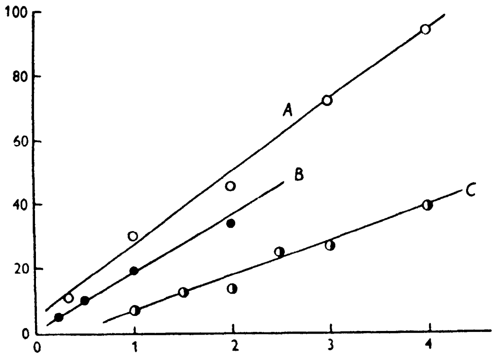


FIG. 58. Relation between dose and volume of juice in three experiments. Line *A* was obtained by using a sample of secretin A, lines *B* and *C* by using an acetone-dried powder D 4. Ordinates: drops of pancreatic juice. For line *A* the abscissae are mg. sample A. For lines *B* and *C* the abscissae are ml. extract, 1 ml. containing the activity of 40 mg. powder D 4. (Burn and Holton, *J. Physiol.* 1948, **107**, 450.)

creatic juice. The enzyme content of pancreatic juice is determined most easily by using Wohlgemuth's method for amylase. The following description is taken from Cole (1941).

Number a series of test-tubes 1 to 10. Into 1 and 2 put 1 ml. of the juice to be tested. Into 2 to 10 put 1 ml. distilled water. Mix the contents of 2, and transfer 1 ml. from 2 to 3. Mix the contents of 3 and transfer 1 ml. from 3 to 4. Continue in this way, rejecting 1 ml. of the fluid from 10.

Prepare starch solution by treating 100 ml. of 2 per cent. soluble starch with 25 ml. of a buffer solution of pH 6·7, 25 ml. 1 per cent. NaCl, and 50 ml. distilled water.

To each tube add 5 ml. 1 per cent. soluble starch. Mix the contents of the tubes and place them in a water-bath at 40° C.,

noting the time. Remove them 30 minutes later, and immerse them in cold water. Arrange them in order in a rack. Add distilled water to near the top of the tube, and 3 drops of N/50 iodine, and mix. Tube 1 will probably give no colour, while tube 10 will give a deep blue. Between these limits will be tubes which are red or purple. Select the tube with the lowest number that has a blue tinge mixed with the red. In the tube numbered

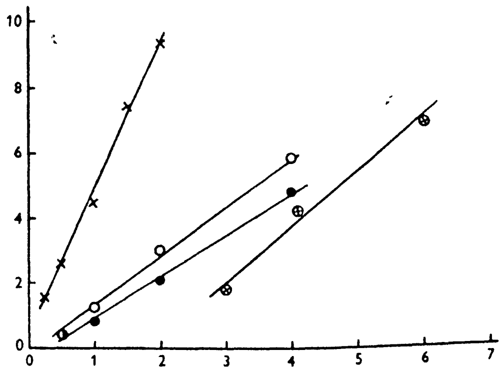


FIG. 59. The relation between dose of pancreozymin and amount of amylase secreted in four experiments. Ordinates: volume × log diastatic power. Abscissae: ml. extract of D 3 injected. (1 ml. corresponds to 40 mg. powder.) (Burn and Holton, *J. Physiol.* 1948, **107**, 452.)

one below this, there was enough enzyme to convert 5 ml. soluble starch to erythrodextrin in 30 minutes.

Wohlgemuth indicates the concentration of enzyme in the following way. He determines the volume of 1 per cent. starch that would be converted to erythrodextrin by 1 ml. of the enzyme solution in 30 minutes. This is the diastatic power, *D*, of the solution. Thus in the experiment if the sixth tube were blue-violet and the fifth tube were red, then, since the fifth tube contains 1/16 ml. of the juice and this amount converted 5 ml. of 1 per cent. starch in 30 minutes, 1 ml. would convert 80 ml. If the temperature of the bath was 40° C., then, since the time was 30 minutes, $D\frac{40°}{30'} = 80$.

In comparing two samples of pancreozymin, it is first necessary to make sure that the juice secreted in response to an

injection of secretin has little or no enzyme content as shown by a diastatic power of 20 or less. The juice secreted by some cats in response to the first or even second injection of secretin contains some enzyme, but after the second injection there is usually no more enzyme, and the comparison of the pancreozymin solutions can then begin. Different doses of one extract are then injected; the volume of the juice secreted in response to each dose and the diastatic power of that juice are determined. It has been found, as shown in Fig. 59, that there is a linear relation between the amount of pancreozymin injected and the product of the volume and log diastatic power. The line giving this relation differs for different cats. The relation between dose and the product 'volume × log diastatic power' is thus determined for one of the samples, and plotted graphically.

Then the effect, similarly measured, of one or two doses of the second sample is determined: from the graph it is possible to see what dose of the first sample would produce the same effect as the injected dose of the second sample, and thus their relative potency is determined.

REFERENCES

Burn and Holton, 1948, *J. Physiol.* **107**, 449.

Cole, 1941, *Practical Physiological Chemistry*, 9th ed., p. 212. Cambridge, W. Heffer & Sons, Ltd.

Harper and Raper, 1943, *J. Physiol.* **102**, 115.

CHAPTER XIX

QUINIDINE SUBSTITUTES

Isolated rabbit auricles. DAWES (1946) has described a method of examining substances which act like quinidine on the refractory period of cardiac muscle. He uses the auricles carefully dissected free from all other tissue taken from the heart of a freshly killed rabbit. When auricles are suspended in a bath of Ringer-Locke solution at 29° C. (not 37°), they soon begin to contract spontaneously provided that the bath is well oxygenated. Some practice and patience is needed to make the preparation successfully. Often an auricle which scarcely beats when first suspended beats well after being left alone for 1 hour.

The two electrodes which are used to stimulate the auricles run down as platinum wires inside a rod of 'Perspex'. The sharpened tips of the electrodes project into a tiny chamber cut out of the bottom of the rod (see Fig. 60). The rod has a central hole, and through this hole a thread, tied on to one auricle tip, is passed until the auricle tip is drawn into the chamber and the electrodes penetrate the substance of the auricle. The chamber at the bottom of the rod has a small oval opening, and when the auricle is drawn up into the chamber the auricle itself seals the opening. A gentle current of warm air is blown down the central hole in the rod and serves the double purpose of preventing Ringer-Locke entering the chamber and so short-circuiting the electrodes, and of oxygenating the tissue in the chamber. Thus the main part of the auricles are in the bath, but the electrodes, being just above the surface of the bath, are outside it.

A thread tied on the opposite tip of the auricle passes round a vulcanite pulley which is immersed in the bath and goes up to a lever working against a light spring. The pulley wheel is loosely fitted on to a glass rod as shown in Fig. 60. The contractions of the auricle are recorded on a smoked drum; they contract spontaneously at a rate varying from 80 to 120 per minute and they can also be stimulated by break shocks from an induction coil at any desired speed, using, for example, a Lewis rotary contact-breaker (C. F. Palmer (London) Ltd., England). The position of the secondary coil is adjusted so that

the peak voltage in the secondary is about ten times that necessary to cause extra systoles at the beginning of the experiment; this ensures that stimuli are so far above threshold that the irregularity of induction shocks will not matter.

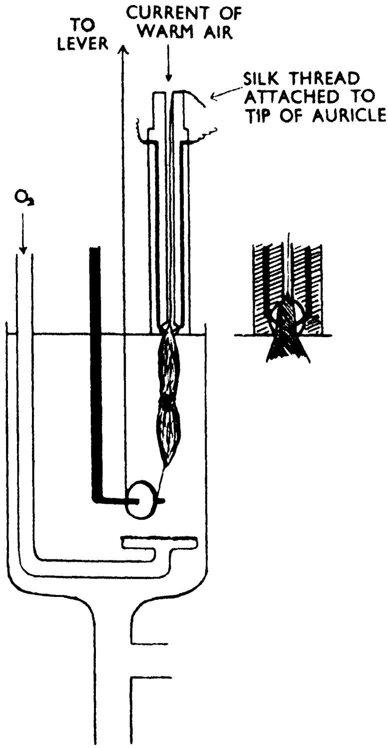


FIG. 60. Preparation of rabbit auricle suspended in oxygenated Ringer-Locke at 29° C. The electrode holder is made of 'Perspex', and the inset shows a closer view of its lower end. (Dawes, *Brit. J. Pharmacol.* 1946, **1**, 90.)

As the rate of electrical stimulation is increased, the auricle follows each stimulus up to a certain point (between 250 and 350 per minute) at which it begins to drop beats (because the interval between shocks is less than the absolute refractory period; cf. Mines, 1913). It is easy to distinguish these dropped beats since the next auricular contraction is more powerful. Thus in the experiment shown in Fig. 61 the auricle followed each stimulus at 200 and 260 per minute; at 266 per minute it

dropped a single beat, and at 274 per minute the response soon became very irregular; at 338 per minute it had adopted a 2:1 rhythm. In this instance the maximal rate at which the auricle could respond would be recorded as 260 per minute. The method is based upon the observation that quinidine reduces this maximal rate.

The maximal rate is measured by applying stimuli at varying

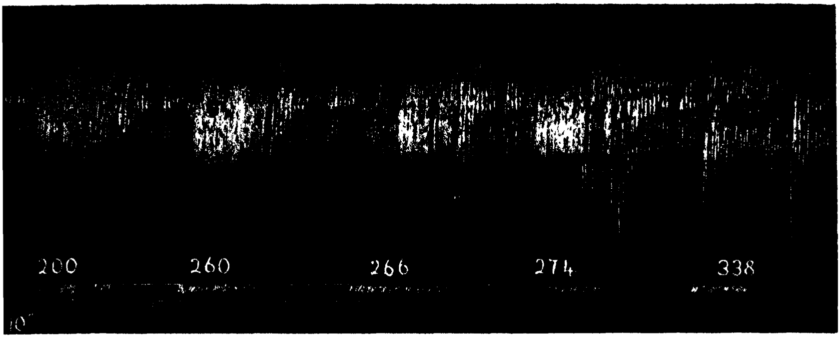


FIG. 61. Record obtained from preparation shown in Fig. 60. The upper signal-marker indicates the duration of stimulation by break induction shocks; the number of stimuli per minute is recorded below this. Up to 260/min. the auricle responds to each stimulus; at 266 and 274/min. it just fails to follow; at 338/min. it adopts a 2:1 rhythm. (Dawes, *Brit. J. Pharmacol.* 1946, **1**, 91.)

rates, each for 10–15 seconds (the maximal rate remains steady over long periods of time, provided no drug is added to the bath). Quinidine is then added and the maximal rate measured again at the end of 10 minutes. Fig. 62 illustrates this procedure. Before the quinidine was added to the bath the maximal rate at which the auricles would follow electrical stimulation was 218 per minute; 10 minutes after the addition of quinidine, the auricles followed stimuli at 184 per minute, but not at 188 per minute; i.e. the maximal rate was reduced by 34 per minute or 15·6 per cent. Preliminary experiments with quinine and quinidine showed that the percentage reduction in the maximal rate bore a linear relationship to the logarithm of the concentration. As Fig. 63 shows, the slope of this line for various drugs is sufficiently parallel to allow their relative activity to be expressed in a single figure. Quinidine is used as the standard for comparison on each preparation.

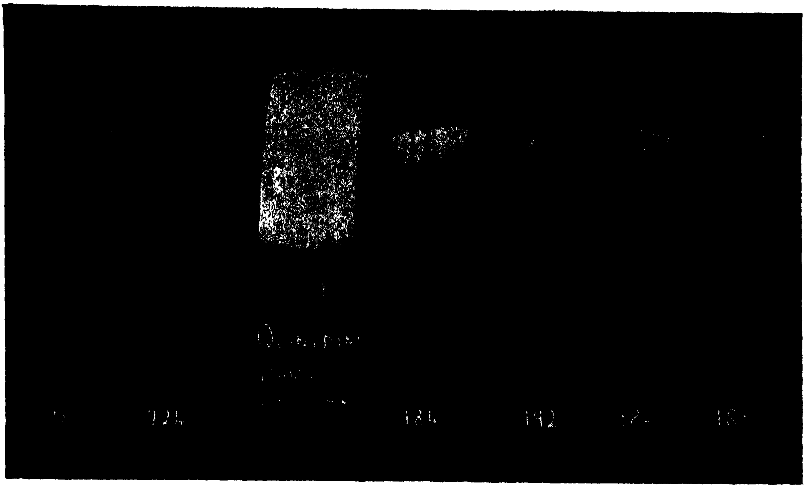


Fig. 62. Preparation as in Fig. 61. The maximal rate at which the auricle responds to electrical stimuli is 218/min.; after quinidine 1 : 100,000 for 10 minutes this is reduced to 184/min. (The smoked drum was stopped for 9 minutes at the arrow, and was run at reduced speed immediately before and after it.) (Dawes, *Brit. J. Pharmacol.* 1946, **1**, 91.)

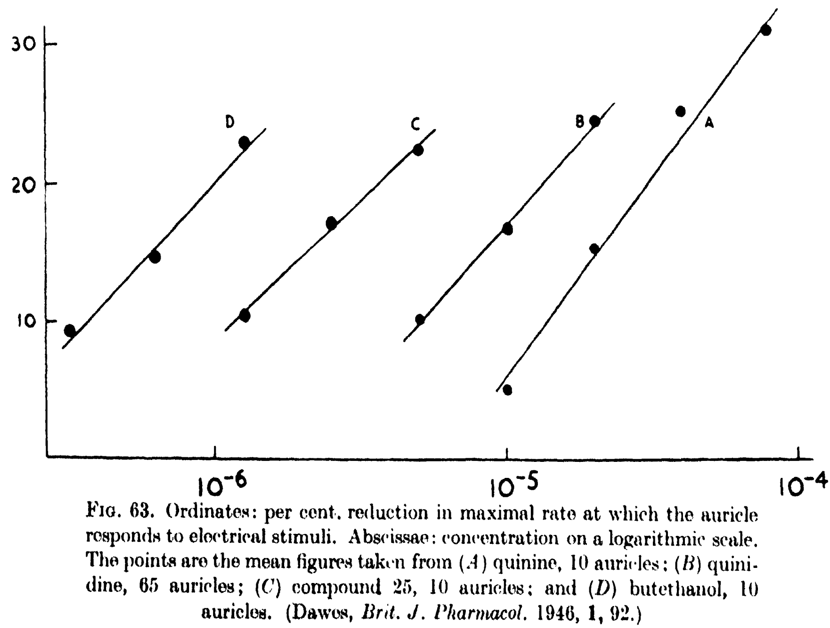


Fig. 63. Ordinates: per cent. reduction in maximal rate at which the auricle responds to electrical stimuli. Abscissae: concentration on a logarithmic scale. The points are the mean figures taken from (*A*) quinine, 10 auricles; (*B*) quinidine, 65 auricles; (*C*) compound 25, 10 auricles; and (*D*) butethanol, 10 auricles. (Dawes, *Brit. J. Pharmacol.* 1946, **1**, 92.)

The auricle returns to its initial maximum rate after the use of quinidine and closely related compounds rather slowly; some cumulation of effect inevitably occurs during the course of the assay, although an interval of 45–60 minutes is normally allowed between each determination. This cumulation does not interfere seriously with the assay, since the percentage reduction of the maximal rate caused by a given dose is, within wide limits, independent of the initial maximal rate. However, in order to reduce this source of error to a minimum it is expedient to use concentrations of quinidine (or of quinidine substitutes) which cause not less than 10 per cent. and not more than 30 per cent. reduction of the maximal rate.

Some idea of the error of the method may be obtained by a consideration of the standard deviations recorded in Table LXXII for those compounds which were tested upon five or more auricles each.

TABLE LXXII

Compound	*Activity in relation to quinidine = 1*	*No. of auricles*
Butethanol	13·8 ± 4·6	10
F1262	4·7 ± 0·67	5
$Ph_2C(OH)COOCH_2N(CHMe_2)_2$.	6·8 ± 1·1	6

REFERENCES

DAWES, 1946, *Brit. J. Pharmacol.* **1**, 90.
MINES, 1913, *J. Physiol*, **46**, 349.

CHAPTER XX

CURARE-LIKE COMPOUNDS

The rabbit head-drop method. THIS method was introduced by H. A. Holaday in 1941, but a detailed procedure was first described by Varney, Linegar, and Holaday (1948). The following account is based on the work of Dutta and MacIntosh at the National Institute for Medical Research in London.

A solution of the drug to be tested is infused at a steady slow rate into the ear-vein of a rabbit until the animal's neck-muscles are so weakened that it can no longer hold up its head, and the amount of drug which must be injected to produce the end-point is noted. The assay is carried out as a cross-over test on 8 rabbits, the unknown and a standard preparation being each administered on the 2 days. Dutta and MacIntosh used an infusion apparatus which enabled 4 rabbits to be treated simultaneously.

Four 5-c.c. syringes are arranged in parallel, each syringe being connected by a tube with a needle which is inserted into the ear-vein of one rabbit and held in place by a piece of adhesive tape. The pistons of the four syringes are driven inwards by a slowly advancing rod which is connected to all four pistons. This rod is driven forward at constant speed by a gramophone motor suitably geared. The pistons of the syringes are lubricated with liquid paraffin. Syringes of nearly the same diameter are chosen and the volume of solution delivered by each for each rotation of the motor spindle is determined gravimetrically with the aid of a revolution counter. The infusion rate was about 0·4 c.c. per minute, and using 2-kg. rabbits the end-point was reached in about 10 minutes.

The rabbits sit side by side in wooden stalls with their heads projecting beyond the edge of the stall, and the syringes are behind and above the rabbits. Each rabbit occupies the same stall on the 2 days of the test and receives the drug from the same syringe. The needle in his ear-vein is connected by polythine tubing to the syringe and is held in place by a bulldog clip fixed just peripheral to the puncture; a strip of adhesive tape binds the plastic tubing to the ear. The plastic tubing is preferable to

rubber tubing because it scarcely alters its capacity when coiled or extended. 'Portex' tubing of 2-mm. bore and buttless needles ($\frac{7}{8}$ in.) of the Solila pattern are satisfactory.

Procedure. The rabbits having sat for 10 minutes or more in their stalls, the apparatus is set running and the needles are successively inserted. The operator calls out the moment of insertion, and an assistant notes the reading of the counter. The rabbit looks normal until about two-thirds of the head-drop dose has been given. The onset of curarization is usually heralded by a period of restlessness; the pupils may dilate; and the rabbit's head begins to fall, but can be jerked up again. When the true end-point is reached the neck is fully stretched and toneless, and there is no movement of the head in response to a light touch on the snout. At this moment—it can usually be judged within 10 seconds or so—the needle is pulled out, and the counter reading is noted for a second time. From the difference between the counter readings and the syringe calibration the volume of solution injected is calculated.

The rabbits soon recover and seem quite normal 15 minutes later. Occasionally animals, especially those unaccustomed to the procedure, suffer severe respiratory embarrassment: they recover quickly if a solution of prostigmine and atropine (prostigmine methylsulphate 0·5 mg., atropine sulphate 0·5 mg.) is promptly injected, through the original puncture site, into the vein. It is very important not to excite the rabbits when they are becoming curarized, or after the end-point is reached, because any struggling deepens the curarization. For this reason docile animals should be selected.

Factors influencing the accuracy of the result. With rabbits which are accustomed to the procedure there is usually no significant increase or decrease in sensitivity with repetition of the treatment. Untrained rabbits, or those which have not been used for some weeks, are apt to be unusually resistant to tubocurarine, but a single test raises their sensitivity to a roughly steady level. It is best if all the rabbits in a cross-over test have been used already within the preceding week: if fresh rabbits have to be used, they should be divided equally between the two groups. Variance between rabbits is highly significant, and the cross-over design is therefore necessary for maximum efficiency.

Variations in the rate of administering the drug have little effect on the dose required to curarize over the range of rates in this test, as is shown by the insignificance of variance between treatments. The average time required to curarize, in this test, varied between 7 minutes for the strongest solution and 12 minutes for the weakest solution: and within these limits differences in the mean duration of the infusion as between standard and unknown solution can probably be disregarded with safety. With infusion rates much slower or faster than these the amount of tubocurarine required to produce the end-point is somewhat increased with a slow injection because an appreciable part of the dose has been inactivated before the injection is completed, and with a fast injection because the drug does not become fully effective immediately on entering the ear-vein, so that the curarization continues to get deeper after the end of the infusion.

The amount of the standard and the amount of the unknown required to curarize each rabbit is calculated in mg. or c.c. per kg. Then the mean figure for the standard and the mean figure for the unknown are equated. Dutta and MacIntosh say that in eight consecutive assays of known solutions of *d*-tubocurarine, using eight rabbits in each assay, the largest error in estimating the mean potency ratio was 8 per cent.

The rat diaphragm method. There are two methods for estimating substances with curariform activity in which an isolated muscle preparation is used. One is the frog *rectus abdominis* preparation, in which contractions are obtained by adding acetylcholine to the bath. A series of contractions to the same dose of acetylcholine are obtained. When *d*-tubocurarine is first added and left to act for a fixed time before the acetylcholine, the contraction is reduced to a height which depends on the amount of *d*-tubocurarine. This method of estimation is not specific for a curare-like compound, since other substances such as procaine have a similar action.

The rat phrenic nerve-diaphragm preparation described by Bülbring (1946) can be used for the estimation of specific curariform effects, since it is possible to stimulate the muscle either through the phrenic nerve or directly. A diminution in muscle contractions which occurs when a substance is added

to the bath can then be shown to be due to a block at the nerve-ending if the response to direct stimulation remains undiminished.

Adult rats are used. The rat is killed by a blow on the head and is bled. After removing the skin over the chest the thorax is opened along the right side of the sternum and the frontal part of the right thoracic wall is removed. The mediastinum behind the sternum is severed and a cut is made above the frontal insertion of the diaphragm, taking care not to damage the phrenic nerve, which sometimes attaches itself to the ribs. The frontal part of the left thoracic wall is then removed and the phrenic nerve is seen. Both left lobes of the lung are removed. The left abdominal muscles are cut along the costal margin and the last rib is held with a pair of forceps. A strip of diaphragm is now cut out. Two converging cuts are made through the ribs towards the tendinous part of the diaphragm parallel to its muscle-fibres, 3 mm. to the right and 3 mm. to the left of the point where the phrenic nerve enters the diaphragm. The strip is cut out beyond the tendinous part with about 2·5 cm. of phrenic nerve attached to it. The preparation has a fan-like shape, being 3 mm. wide at the tendinous end, while at the costal margin it is about 12 mm. wide. The preparation is fixed by a glass rod with a pin pushed through an intercostal space holding it at the bottom of the organ bath. A thread tied around the tendinous end is attached to a light isotonic heart lever writing on a smoked drum. The nerve is stimulated by a pair of electrodes of the following design. Two solid platinum blocks are each attached to a wire and these are embedded in 'Perspex' (a plastic). The wires run in the handle 10 cm. long, and are bent at the bottom at a right angle. The platinum blocks are thus in the short foot of the ⅃ as shown in Fig. 64. When the 'Perspex' is set, a hole 0·75 mm. diameter is drilled through the foot containing the blocks to form a channel for the nerve, which thus passes through two platinum rings and nearly fills the channel. The electrodes are fixed vertically under the surface of the solution inside the bath. The nerve must be left slack to allow for the movement caused by the contractions. Single square wave shocks are applied to the nerve. Constant muscle contractions can be obtained for many hours in response to single shocks at rates not exceeding 12 per

minute. With faster rates fatigue sets in if they are applied for prolonged periods.

Tyrode solution

NaCl	. 8·0 gm.
KCl	. 0·2 ,,
$CaCl_2$	. 0·2 ,,
$MgCl_2$	. 0·01 ,,
NaH_2PO_4	. 0·05 ,,
Dextrose	. 1·0 ,,
$NaHCO_3$	. 1·0 ,,
Distilled water . . .	. 1 litre.

Tyrode's solution containing double the usual amount of dextrose is used. A sintered glass gas-distribution tube is fixed at the bottom of the bath providing vigorous oxygenation with a mixture of 95 per cent. O_2 + 5 per cent. CO_2. The capacity of the bath is 100 c.c. and the temperature is kept between 36° and 37° C.

The first account of using the rat diaphragm preparation for the assay of curare-like substances was given by Chou (1947), who showed that the effect of *d*-tubocurarine was greater when the number of stimuli per minute was greater, and also that the effect was greater when 5 per cent. CO_2 was included in the oxygen with which the bath was aerated. An estimation from West's paper (1947) is given in Fig. 65.

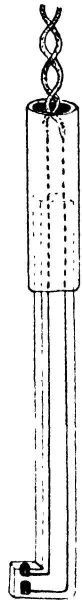


FIG. 64. Electrodes for phrenic nerve (see text). Each platinum block is 1 mm. cube (approx.). The foot is 10 mm. long, 5 mm. high, and 2·5 mm. wide. The handle is 10 cm. long, and 5 × 2·5 mm. in section. (Bülbring, *Brit. J. Pharmacol.* 1946, **1**, 38.)

Trevan has pointed out that the response of the muscle varies with the duration of the stimulus, over a range from 0·33 to 6·0 milliseconds, using a direct current square wave stimulus. When a long duration stimulus was used, the response was larger, and was found to be the resultant of a multiple twitch in the component muscle groups. These responses were found to be more easily blocked by curare than the single twitch responses to short duration stimuli. The two types

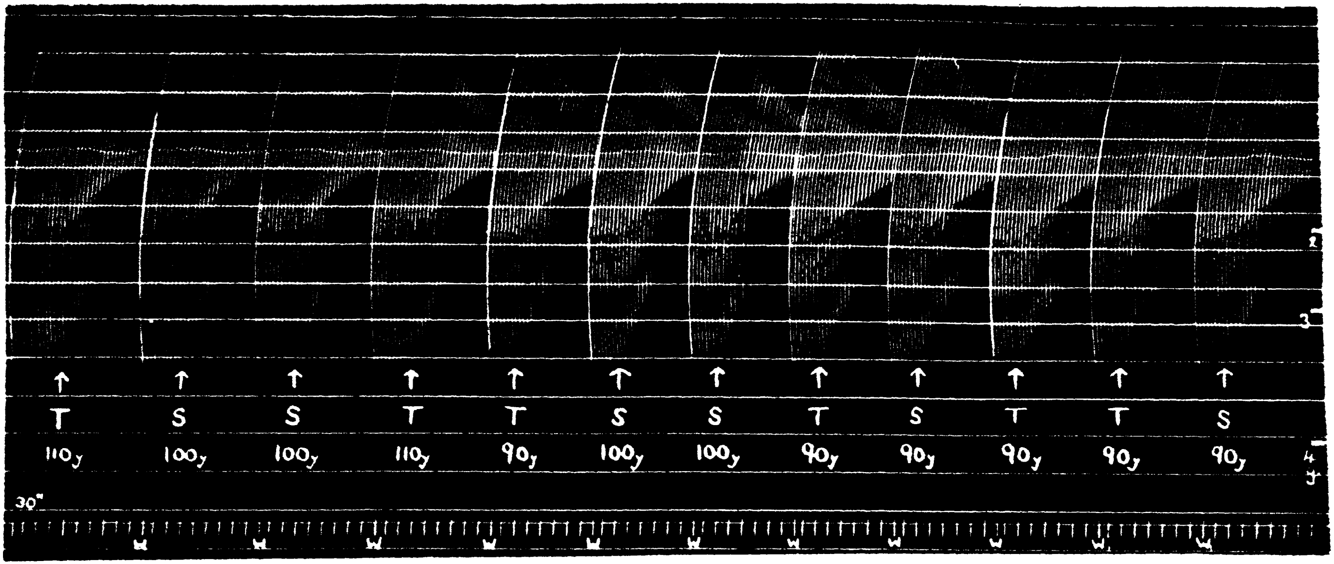


FIG. 65. Illustration of the estimation of an unknown tubocurarine solution (*T*) by comparison with a known solution (*S*).

100 μg. of *S* cause mean percentage inhibition of	58·3
90 μg. ,, ,, ,, ,, ,, ,, ,,	53·8
110 μg. ,, *T* ,, ,, ,, ,, ,,	60·8
100 μg. ,, ,, ,, ,, ,, ,, ,,	54·6
90 μg. ,, ,, ,, ,, ,, ,, ,,	50·8

(West, *Quart. J. Pharm. Pharmacol.* 1947, **20**, 526.)

of response also differed in their constancy, the multiple twitch response showing variations during an experiment, sometimes suffering such reduction that by the end of an experiment they were no greater than the single twitch. The single twitch, however, remained constant in size up to 26 hours after isolation.

Trevan recommended 10-minute intervals between successive doses of drug, the drug being allowed to act for 5 minutes before washing. Chou added doses at 13-minute intervals, allowing 3 minutes for the action of the drug before washing out.

Effect of temperature. West (1947) made an important contribution to the rat diaphragm method by discovering that the preparation maintained the size of contraction originally produced for a longer period if the temperature of the bath was lowered from 37° to 20° C. He found further that the rate of recovery of the muscle after the addition of a dose of *d*-tubocurarine was greatly accelerated by adding KCl to the bath when the *d*-tubocurarine was washed out. To a 75-c.c. bath, 45 mg. KCl was added and stimulation was interrupted for 2 minutes. The result was to make it possible to add doses of *d*-tubocurarine at intervals of 8 minutes (which were allowed to act for 3 minutes) and to have full recovery after washing out each dose. The KCl added was of course removed before adding the next dose of *d*-tubocurarine. West used a stimulation rate of 8 per minute.

Species of animal. Trevan observed that the relative potency of a curare-like compound and *d*-tubocurarine on the phrenic nerve-diaphragm preparation differed according to the species used. Details illustrating this were published by Wien (1948), who found that the triethyliodide of tri(diethylaminoethoxy)-1–2–3 benzene was equal to *d*-tubocurarine on the diaphragm of the rabbit, was one-fifth the activity of *d*-tubocurarine on the diaphragm of the kitten, but was less than one-hundredth of the activity of *d*-tubocurarine on the diaphragm of the rat.

The cat method. The rabbit head-drop method has the advantage of using the whole animal, but it has the disadvantage of being unspecific. Substances which are not curare-like in action will cause a head-drop. This difficulty is overcome by the use of the cat method, in which the respiration can be recorded at the same time as the contractions of the gastrocnemius muscle elicited by stimulation of the sciatic nerve. Paton and

Zaimis (1948) have shown that the compound decamethylene bis-trimethyl ammonium differs from *d*-tubocurarine in having much less action on respiration relative to its action on other muscles. By using the cat method, it is possible to compare a substance with *d*-tubocurarine to observe whether doses which are equally active on the gastrocnemius differ in their action on the diaphragm.

The cat is briefly anaesthetized with ethyl chloride followed by ether; when a trachea tube has been inserted, a cannula is tied in the right external jugular vein and a solution of chloralose (80 mg. per kg.) is injected.

A steel rod is then fixed through the lower end of the femur of one hind-leg. This is done by first making a skin incision on the outer side and then drilling a hole with a small bradawl through the femur. Successively larger sizes of bradawl are passed through the hole until the steel rod can be driven through. The ends of the steel rod are then fixed rigidly in clamps so that the femur rises vertically from the table. After excising the skin covering the heel, the piece of bone into which the gastrocnemius is inserted is cut through with bone forceps, and strong wire is then wrapped round the gastrocnemius tendon. The wire is attached to a tension lever, the writing-point of which records on a smoked drum. By an incision through the skin covering the back of the thigh the sciatic nerve is exposed between the hamstring muscles and crushed proximally between artery forceps, though the nerve is not divided. Shielded electrodes are then placed around the nerve so that stimulation can be applied at a rate which may be from 8 to 20 per minute. Square wave stimuli of short duration (not more than 1 m. sec.) should be used.

Injections should be made into the artery leading to the leg so that the doses of the curarizing agent can be as small as possible. If the gastrocnemius of the left hind-leg is prepared, a cannula can be placed in the right external iliac artery pointing towards the bifurcation of the aorta, so that when an injection is made the material passes into the blood-stream at the bifurcation and is carried down the left femoral artery. The effect of injections is illustrated in Fig. 66, taken from Bülbring and Chou (1947). To make a comparison of two curarizing agents doses which produce effects of the same duration are

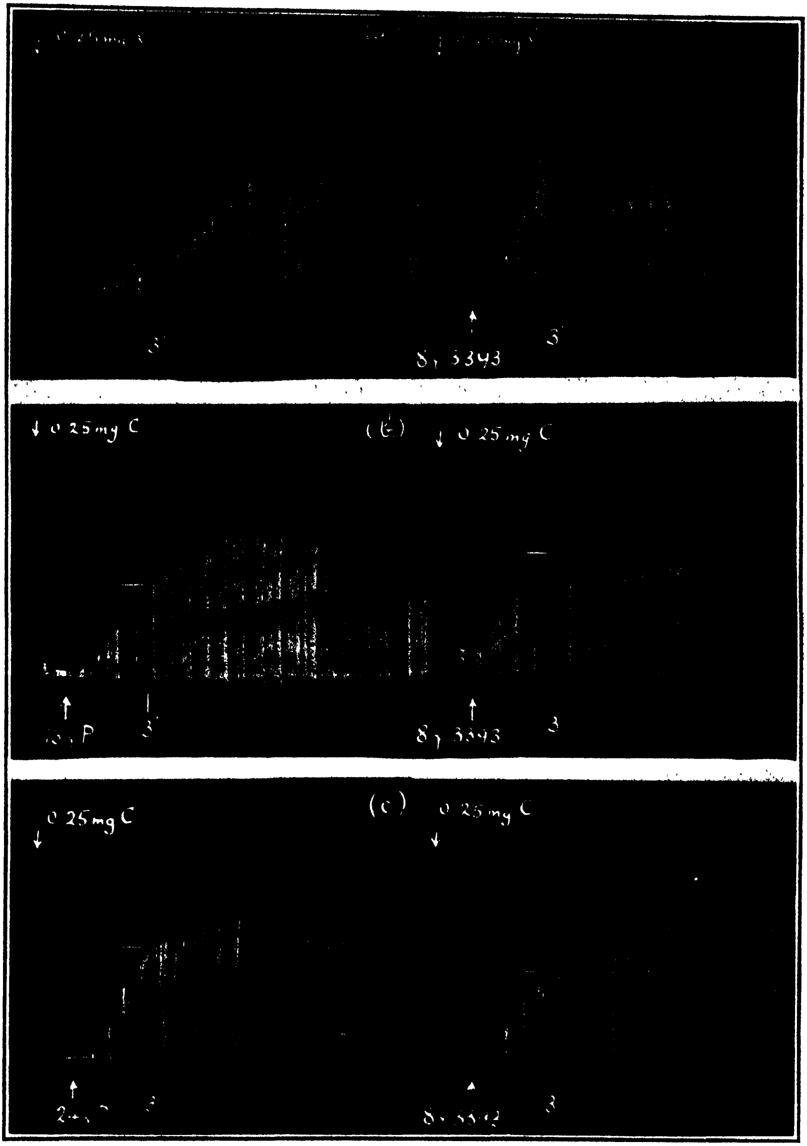


FIG. 66. Cat. Sciatic gastrocnemius (contractions upwards). Rate of stimulation 12 per min. (*a*) Recovery from the effect of 0·25 mg. tubocurarine compared with that after 8 μg. 3393. (*b*) The same after 16 μg. prostigmine and 8 μg. 3393. (*c*) After 24 μg. prostigmine and 8 μg. 3393. (Bülbring and Chou, *Brit. J. Pharmacol.* 1947, **2**, 16.)

determined. In starting an experiment a resting tension of about 1 kg. is applied to the gastrocnemius.

The respiration can be recorded at the same time by any suitable method. That described by Gaddum (1941) is very convenient since it affords a measure of the rate and depth. A short tube in the trachea has valves so that inspired air comes through one valve and expired air leaves by the other. The inspired air comes from a large bottle of 5–10 litre capacity to which air is admitted through a capillary tube passing through the stopper of the bottle. The capillary tube acts as a resistance. When the cat inspires, air is taken out of the bottle faster than it can enter, and the pressure falls. This fall of pressure is recorded on a drum by connecting the large bottle to a tambour with a rubber top of 11 cm. diameter. This rubber top is sucked down when the pressure in the bottle falls and the movement of the rubber is recorded by a writing-lever.

REFERENCES

Bülbring, 1946, *Brit. J. Pharmacol.* **1**, 38.
Bülbring and Chou, 1947, ibid. **2**, 8.
Chou, 1947, ibid. **2**, 1.
Gaddum, 1941, *J. Physiol.* **99**, 257.
Paton and Zaimis, 1948, *Nature*, **162**, 810.
Varney, Linegar, and Holaday, 1948, *Fed. Proc.* **7**, 261.
West, (1947), *Quart. J. Pharm. Pharmacol.* **20**, 518.
Wien, 1948, *J. Physiol.* **107**, 44 P.

CHAPTER XXI

ANTHELMINTHIC SUBSTANCES

LABORATORY experiments upon diseases caused by worms require either a supply of naturally infected animals or a complicated arrangement for the cyclical transmission of the parasite through the appropriate intermediate host. Dependence upon finding a source of animals carrying a natural infection has many drawbacks, and for well-controlled experiments the establishment of a colony of the intermediate host is essential. The maintenance of cyclical transmission requires a great deal of care and attention and, as the development of adult worms in the mammalian host may take many weeks, a large stock of animals must be carried to ensure a regular supply for chemotherapeutic tests.

1. Experimental Filariasis

Culbertson and Rose (1944) showed that the cotton rat, *Sigmodon hispidus*, which is naturally infected with the filarial worm *Litomosoides carinii*, could be used for tests upon the filaricidal action of drugs. Williams and Brown (1945) found the intermediate host to be the rat mite *Liponyssus bacoti*, and since that time colonies of the mites and infected cotton rats have been established in several laboratories in the United States and Great Britain.

The adult worms are colourless thread-like beasts; the males are about 20 mm. long and 0·2 mm. in diameter, and the females are up to 150 mm. long and 0·3 mm. in diameter. They are found in the pleural cavities of the rats. After fertilization, the female worms liberate motile microfilariae which swim in the peripheral blood. Mites feeding upon the blood of infected rats ingest some of the microfilariae, which grow in the bodies of the mites. From the mites they are transmitted to other rats; the exact mode of entry of the worms into the new host is not yet fully understood. It takes about 50 days for the worms to mature in a new host.

The techniques described briefly below are those used at the National Institute for Medical Research at Hampstead, and

are described in greater detail by Hawking and Sewell (1948). In this excellent paper all the details and snags of the method are discussed, and full references to the literature are given.

Cotton rats. Cotton rats are not very easy to breed in the laboratory because of the low fertility-rate, and as they are very susceptible to many bacterial and virus infections great attention must be paid to the cleanliness and general hygiene of the colony. Cotton rats are more difficult to handle than ordinary rats, because they are very swift-moving, and are timid, and bite when they are frightened. They should be grasped firmly by the scruff of the neck and never handled by the tail. The skin of the tail is thin and is very readily stripped off. Cotton rats are fed upon ordinary rat cubes, supplemented with green vegetables. Females may be allowed to run in cages together, but the males must be kept separately or they fight and kill each other.

Maintenance of a colony of mites. Mites are kept in an insectarium similar to that used for the breeding of mosquitoes (Bertram and Gordon, 1939; Hunt and Davey, 1947). The temperature must be between 24° and 26° C., and the relative humidity about 80 per cent. Glass jam-jars of the 1-lb. size are cleaned and a 2-inch depth of sawdust covered with a filter-paper is placed in each. The mouths are closed with fine silk or nylon fabric tied down with string, and the jars are sterilized in the autoclave and dried. Adult mites are removed from existing cultures with a fine brush, and cleaned by floating on to the surface of some water. They are transferred to filter-paper and examined with a lens to make sure that they are of the correct species and have no smaller mites crawling on them. About ten mites are transferred to a clean jar, and these will multiply to about 5,000 in the course of a month. They are fed by putting a day-old mouse into the jar every day, or a day-old rat every other day. The baby mice must be freed from foreign insects by dipping them under the surface of some ether for a few seconds, without killing them, and allowing them to dry before putting them into the jars. Dead mice are removed daily. A thin smear of dimethylphthalate around the lip of the jar will prevent the escape of mites when the cover is removed.

The chief danger to the culture is contamination with un-

wanted species of mites which compete with and kill the *Liponyssus*.

Infection of mites with filaria. In order to obtain mites containing worms, cotton rats with microfilariae in the circulating blood are kept in contact with a colony of mites for a week. The technique used is designed to ensure (*a*) that the temperature, humidity, and degree of cleanliness is suitable for the

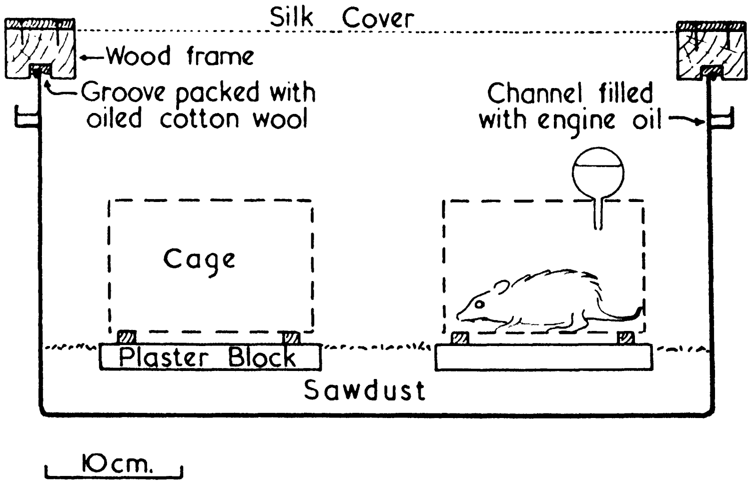


FIG. 67. Tank for the infection of mites and cotton rats with filaria.

breeding and feeding of the mites; (*b*) that the colony remains free from contamination with unwanted mites, lice, and flies.

The mites and rats are kept in rectangular galvanized iron tanks 54 × 50 × 26 cm. in size, fitted with lids made of fine silk fabric stretched on a wooden frame. The lid has a recess packed with cotton-wool soaked in liquid paraffin into which the edge of the tank fits. A channel containing engine-oil surrounds the outside of the tank 4 cm. from the top, in order to prevent the escape of mites (see Fig. 67). The tanks are steam-sterilized; the lids are sprayed with xylene, put on the tanks, and allowed to dry. Sterile sawdust is put into the bottom of each tank to the depth of 5 cm., and upon this are placed blocks of cast plaster of paris, 2 cm. thick and slightly larger than the small cages in which the cotton rats are to be kept. The plaster absorbs urine and prevents waterlogging of the sawdust. The cages are

supported about 0·5 cm. above the blocks to prevent caking of faeces upon the wire bottoms of the cages.

Cotton rats showing more than 50 microfilariae per μl. of the peripheral blood are selected, and isolated in small sterile cages 15×10×10 cm. in size. The rats are carefully sprayed with xylene and left to dry over a tray of lysol. This is to remove any mites or lice they may be harbouring. Two or three rats in their cages are then put into each prepared tank, and a jam-jar full of rich mite culture is added. The tanks are kept in a room at 24°–26° C. and the humidity is adjusted so that the air inside the tanks is about 80 per cent. saturated. To avoid accumulation of waste food and water, no more than 2–4 gm. of baked cube diet and 10 ml. of water is allowed for each rat daily. The mites multiply rapidly and at the end of a week will be swarming over the sawdust, and a number of them (4 per cent. or more) will be carrying filaria.

During manipulations, the hands of the operator should be smeared with dimethylphthalate.

Infection of clean rats. Within 3 days of removal of the infected rats from the tanks, clean cotton rats in small cages, sprayed with xylene and dried as before, are packed into the tanks and left there for 14 days. Each tank will take 3 cages containing 5 female rats, or 8 individual cages containing male rats. Food and water is restricted and urine is absorbed on plaster blocks as before. At the end of the fortnight the rats are removed, sprayed with xylene to kill the mites, and stocked in separate cages for 6 weeks for the worms to reach maturity.

Examination of infected rats. At the end of the 6 weeks, the blood of each rat is examined for microfilariae; 0·05 ml. of blood is taken from the tail into a graduated Pasteur pipette and spread on a slide to make a thick film. The film is stained with Giemsa and the number of microfilariae present is counted. Most of the rats will be found to have 10–15 parasites per μl. of blood, and these are suitable for chemotherapeutic tests. Rats found to be too lightly infected should be kept a little longer for more microfilariae to accumulate in the blood, and used for the infection of further tanks of mites.

Treatment with drugs and assessment of effect. The most economic procedure for a screening test for new drugs is to give

doses daily for 6 consecutive days. Doses may be given intraperitoneally, and should be chosen to be as close as possible to the toxic range. Hawking and Sewell (1948) found that a suitable dose-level was 0·4 times the dose which just failed to kill mice when given intraperitoneally on 4 successive days. The use of mice for a rough toxicity test conserves cotton rats, which are usually not so cheap or so plentiful as mice. It is important to use sterile solutions of drugs, and to take aseptic precautions for the injections, because of the susceptibility of cotton rats to secondary infections.

Six days after the last dose of drug, the blood of each rat is again examined for microfilariae, and the rats are killed. The thorax is opened and all of the worms found in the pleural cavities are removed. Some worms will be obviously living and motile; others may be stimulated to activity by suspension in Ringer's solution at 37° C. If aggregated masses of worms are found, they are examined under the microscope and the proportion of living and dead ones determined.

In untreated animals the microfilarial count in the peripheral blood increases until it reaches a maximum level in about 100 days. If, therefore, the rats are used 6 weeks after exposure to the mites as recommended, any check produced by drug treatment upon the reproductive processes of the adolescent worms will be detected by a lower intensity of microfilarial infection in the blood.

Effective treatment with drugs such as the organic antimonials produces compact masses of dead worms in the pleural cavity, embedded in fibrin. Microfilariae are more resistant to antimony than the adult worms, and frequently persist in the blood-stream for several months after treatment. On the other hand, drugs of the piperazine series are more effective upon the microfilariae than the adult worms. Microfilariae rapidly disappear during treatment, probably because they become entrapped by macrophages in the sinusoids of the liver.

This test has not yet reached the stage at which accurate quantitative assays of drugs may be carried out. Assessment of the potency of a drug must be made after due consideration of the effects produced upon the adult worms and the microfilariae, compared with the conditions in untreated control animals.

2. Experimental Schistosomiasis

Schistosoma mansoni and *japonicum* can be transmitted to mice and other laboratory animals. *S. haematobium* is very difficult to transmit to snails in Great Britain and has not yet been used successfully in the laboratory. Infected animals pass eggs in the faeces. These will hatch out in water under a strong light to liberate miracidia which will infect snails. After about 3 weeks, infected snails begin to liberate cercariae into the surrounding water and these can be used to infect mice. In the mammalian host, the cercariae develop into adult worms which migrate and pair in the mesenteric veins. Eggs are shed and pass through the wall of the gut into the faeces, thus completing the cycle.

Experimental schistosome infections have been used for the testing of chemotherapeutic substances for a number of years. Kikuth at Elberfeld examined a very large number of drugs and discovered the activity of the thioxanthone derivative, miracil. Schubert (1948, 1949) in the United States developed a method and also examined a large number of substances. The methods described below are essentially those used by Schubert, with modifications by Standen.

Suitable hosts. Mice are the most convenient animals for chemotherapeutic tests, and though the number of ova excreted in the faeces is small, mice can be used for complete cyclical transmissions. An infected hamster or monkey is very useful to keep as a source of ova, because these animals discharge them in much larger numbers.

The best snail vectors to use for *S. mansoni* are *Australorbis glabratus* and *Planorbis boissyi*. *A. glabratus* is the easier to rear, and produces larger numbers of cercariae than *P. boissyi*.

Maintenance of the snail colony. Snails may be kept in glass tanks of water aerated by means of a small aquarium pump. The temperature of the water is kept at 25°–28° C., and the snails are fed two or three times a week with scalded lettuce leaves. Such aquaria are simple, but require constant attention and frequent cleaning and provide very artificial conditions for the snails. A little trouble expended at the outset upon the establishment of biologically balanced aquaria as suggested by Standen (1949) is well repaid. A tank of 20–25 litres capacity

is filled with water with a bed of pond-mud and sand. London tap-water is suitable if it is 'conditioned' by allowing it to stand in the laboratory for a few days to remove traces of chlorine or other volatile noxious substances. Water from a copper-piped supply is toxic and must not be used. The mud at the bottom of the tank contains large numbers of *Tubifex* which destroy snail faeces and rotting vegetation. *Daphnia* keeps the water clear by consuming protozoa and algae. Oxygen is supplied by water-plants, the most suitable being *Myriophyllum*, *Ludwigia*, and *Vallisneria*. These are powerful oxygenators, they serve as 'nurseries' for the young snails, and are usually not eaten by the snails if scalded lettuce leaves are offered as an alternative. Predaceous creatures such as tadpoles, fish, leeches, water-mites, and water-boatmen must be eliminated. A well-balanced aquarium will maintain itself for 6 months to a year without cleaning.

Snails breed at temperatures between 22° and 28° C., but infected snails must be kept at 28° C. or the liberation of cercariae is delayed. The snails die if the water becomes acid, but a balanced aquarium will remain at a suitable pH of 7·3 to 8·0 and will give warning of any unexpected changes by variation in the relative populations of its flora and fauna. Additional lighting must be supplied during the winter months if water plants are used as the source of oxygen, though the snails themselves do not require a high light intensity for breeding. 'Warm white' fluorescent lighting is suitable.

Infection of snails. Faeces containing ova of *S. mansoni* are ground gently with fine sand, mixed with about 10–15 times their volume of saline, and shaken for 10 minutes. The suspension is filtered through gauze and allowed to sediment for 15–20 minutes. The supernatant fluid is rejected, the sediment shaken with a further quantity of saline and again allowed to stand. The sediment is rinsed into a beaker with boiled water at 28° C., and placed in a bright light. In ½–1 hour, miracidia begin to hatch out from the ova, and if the beaker is held in a strong beam of light from a 'Pointolite' bulb and condenser, they may be seen swimming about vigorously. They may be picked up with a capillary pipette. Young adult snails, 2–3 months old, are placed individually in 5-ml. beakers, each containing 2 ml. of water, and 6 miracidia are added to each.

They are allowed to stand at 28° C. overnight to provide ample time for the miracidia to penetrate the bodies of the snails. Less than 4 miracidia per snail may give failures of infection, and more than 8 may cause early death. The snails are transferred to prepared aquaria, and kept at 28° C. They begin to produce cercariae after 3 weeks and continue to do so every day until they die 8–10 weeks later. To test for cercarial production, snails are put into glass tubes (3 in. × 1 in.) containing water at 9 a.m. and left under a bright light until 1 p.m. Most cercariae emerge during the morning; only small numbers emerge in the afternoon. The presence of cercariae is readily detected by holding the tube in a beam of light and examining with a hand lens. Snails which are producing cercariae are stored in a prepared aquarium until required for the infection of animals. They are then again put into tubes of water in the light and the cercariae which emerge are taken up and counted into a syringe fitted with a wide-bore needle. Pooling of cercariae from several snails is important or unisexual infections may result.

If *P. boissyi* is used, up to 40 per cent. of the snails may die during the incubation period; only about 12 per cent. of the snails infected produce cercariae. The mortality rate of *A. glabratus* is much lower.

Infection of mice. The mice used for infection should be from a good clean stock and free from other infestations such as *Cysticercus fasciolaris*. The animals need to be stocked for about 12 weeks, and unhealthy specimens will not survive so long.

Mice may be infected either by injecting the cercariae with a syringe, or by allowing them to penetrate the skin. Schubert found it most convenient to inject about 150 cercariae intraperitoneally to each mouse. Standen (1949) finds that cutaneous penetration gives a greater proportion of successful infections than parenteral injection. The method used is first to let the mice paddle for 5 minutes in a tray containing about ¼ inch depth of warm water, to wash their feet and to encourage the voiding of urine and faeces before putting into water with the cercariae. Each mouse is then put into a small glass jar, 60-ml. capacity, and water at 25° C. is added. The jar is small enough to restrict movement and is fitted with a bung with an air-hole as in the diagram (Fig. 68). About 100 cercariae (collected from

many snails and pooled to minimize the chance of unisexual infection) are then added to the water in the jar through the air vent and the mouse is left to bathe for about ½ hour at 25° C. The cercariae penetrate the skin and disappear. Great care must be taken to avoid contact of water containing cercariae with the skin of the operator. If such contact is suspected, the skin is at once scrubbed with alcohol. The mice are put into a cage of warm hay to dry, and are then stocked for at least 50 days. At the end of this time schistosome ova may be found in the faeces when a fresh smear is examined under the microscope. All of the animals usually become infected, though the number of worms found in each will vary considerably. Standen found the average number per mouse to be 18.

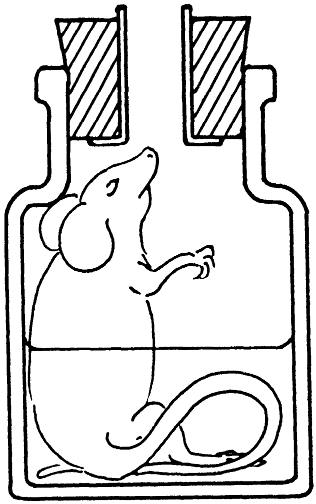


FIG. 68. Bottle for infection of mice with schistosome cercariae.

Treatment with drugs and assessment of effect. Untreated animals show enlargement of the liver and spleen 50 days after infection. Pairs of worms *in copula* are found in the mesenteric blood-vessels and worms are also present in the portal vein. At a later stage the liver and spleen become grossly enlarged and covered with small white granules; these are areas of tissue reaction surrounding ova, some of which contain living miracidia, and some of which are dead. Schubert (1948) found that if active drugs were given, the infection was not always eradicated, but the worms migrated from the mesenteric veins into the liver, and he based his method of assay upon the extent of this migration. After a few weeks, the worms began to return to the mesentery. He therefore allowed 2 weeks after conclusion of treatment, and then killed the mice and counted the worms. If a large dose of an active drug is given by stomach tube twice daily for 5 days, a period of 1 week after the last dose is often sufficient to demonstrate an effect upon the worms.

Each mouse is killed by dislocating the neck, and the abdomen is opened. The portal vein is ligatured with fine thread close to

the liver and the mesentery spread out and examined from one end to the other. Paired worms are removed with forceps and placed in saline at 37° C. to test for viability, and counted. The portal vein is then cut across just distal to the ligature and squeezed gently in saline to expel the worms, which are also counted. The liver is then removed in one piece and is crushed firmly between two 6×6-inch pieces of plate glass. When held up to the light the worms present in the liver can easily be counted. The numbers of worms of each sex in each location are noted, and the distribution in treated animals compared with that in untreated controls infected at the same time.

The male worms are about 6·5–9 mm. long; the females are 7·2–14 mm., and are thinner and more heavily pigmented than the males. There are always more males than females and solitary males do not usually reach the mesentery. Therefore, a count based on the total number of worms of both sexes is not reliable as an indication of retreat towards the liver. It is much better to follow the movement of the female worms, because these are normally all found in the mesentery, paired with the males, and do not leave this situation except under the influence of drug treatment.

Table LXXIII shows the distribution of worms in two groups of mice with 70–77-day-old infections. One group was treated with nilodin twice daily by mouth. The mice were killed 7 days after the last dose. The worms in the treated mice had migrated from the mesenteric veins to the liver.

TABLE LXXIII

Treatment	*No. of mice*	*Mean no. of worms per mouse*	*Per cent. of ♂*	*Per cent. of ♀*	*Per cent. paired*	*Distribution*		
						Per cent. in liver	*Per cent. in portal vein*	*Per cent. in mesenteric veins*
Controls . .	15	16	58	42	80	21	22	57
Nilodin (miracil D) 50 mg./kg. 10 doses .	9	17	64	36	50	97	3	0

Estimation of anticestode drugs

Substances which are of use against tapeworms can be tested by using mice infected with *Hymenolepis diminuta*. Mice in England are often naturally infected with the two species,

H. diminuta and *nana*; the presence of the infestation is detected by examining the faeces.

Method of testing (Holton, 1947)

Examination of mice. The mice are fed at least 1 hour before examination. A pellet of faeces is taken from each mouse, and

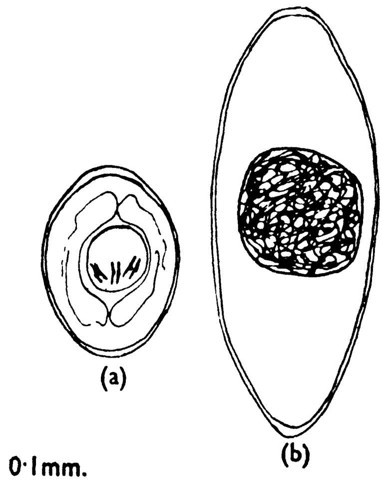


FIG. 69. Helminth eggs found in mouse faeces. (*a*) *Hymenolepis* egg containing a hexacanth embryo. (*b*) Nematode egg. This is the only common egg likely to be mistaken for *Hymenolepis*. (Holton, *Brit. J. Pharmacol.* 1947, **2**, 101.)

a smear is made on a microscope slide. The typical egg (see Fig. 69) containing a hexacanth embryo can then be seen usually with the low power, but with the high power if necessary. It is not possible to distinguish between the two species by examination of the eggs. Of the mice coming to a laboratory from dealers, the proportion infested depends on the dirtiness of the stock. If non-infested mice are mixed with infested mice, a percentage of them become infested in 3 weeks.

Preparation of material. In using Extractum Filicis B.P., the extract is stirred, and about 4 drops are put into a 50 ml. conical flask and weighed. (The weight is about 200–250 mg.) This extract is then dissolved in 0·5 ml. alcohol. Distilled water containing about 1 per cent. sodium glycocholate is added so

that 1 ml. of the resulting emulsion contains exactly 10 mg. Ext. Filicis. The emulsion is shaken before use and diluted as necessary with a solution of sodium glycocholate and alcohol in the same proportions as just given so that a volume of 1 ml. is administered to each mouse.

Performance of test. To prepare mice for the test, they are kept in boxes containing lumps of sugar overnight and are given water. Thus by the morning the intestines are empty of solid food and yet the mice are not weakened as they would be if kept without food. Each of a group of (say) five mice is given a dose of 1 ml. of the preparation to be tested. This is given by mouth, using a blunted hypodermic needle on the end of a 1-ml. syringe. The needle is passed down into the stomach. The mouse is held by the skin of the back of the neck between the finger and thumb of the left hand, and the tail is held down by the third and fourth fingers against the base of the thumb. To pass the needle the mouse is held vertically.

Each group of mice is placed in a box with a metal floor and is supplied with lumps of sugar and water. After 2 hours each mouse is given 1 ml. warm 6·5 per cent. magnesium sulphate solution and replaced in its box. Three hours after this the examination of the faeces on the floor of the box, and of the intestines, is carried out.

Examination of worms. Worms are divided into four groups:

(*a*) worms in faeces;
(*b*) worms in large intestine;
(*c*) dead worms in small intestine;
(*d*) live worms in small intestine.

Examination of faeces is carried out in water in a black developing-dish. The white fragments of worms are transferred from the box with forceps and rinsed free of debris. The fragments are collected in a small weighed Gooch crucible.

Since tapeworms are normally not found in the large intestine, it follows that worms present there must be included in the category of worms affected by the drug. The large intestine of each mouse is therefore put in saline at 38° C. in a black dish, and slit open longitudinally with scissors. The mucosa is scraped with a microscope slide and the white fragments of worms are transferred to the Gooch crucible.

The small intestine is then opened in the same way in warm saline and the worms found are divided into those which are immobile (and which do not react when pinched in forceps) and into those which are alive. The immobile worms are added to the Gooch crucible containing the worms found in the faeces and in the large intestine, while the living worms are weighed separately in a second Gooch crucible. An example of an experiment carried out by Holton (1947) is given in Table LXXIV.

TABLE LXXIV

	Group 1	*Group 2*	*Group 3*	*Group 4*
Dose of Ext. Filicis	*10 mg.*	*5 mg.*	*2·5 mg.*	*1·25 mg.*
Worms in faeces (*a*)	53	235	286	121
Worms in large intestine (*b*)	65	130	115	0
Dead worms in small intestine (*c*)	0	41	0	0
Living worms in small intestine (*d*)	0	75	290	771
Total worms ($a+b+c+d$)	118	481	691	892
Total worms affected by drug ($a+b+c$)	118	406	401	121
Percentage wet weight of worms affected by drug	100	85	58	13

Figures are weights of worms (mg.).

TABLE LXXV. *Percentage wet weight of tapeworms affected by doses of Ext. Filicis given to groups of five mice*

	Group 1	*Group 2*	*Group 3*	*Group 4*
Dose of Ext. Filicis	*10 mg.*	*5 mg.*	*2·5 mg.*	*1·25 mg.*
Experiment 14	100	55	48	0
15	91	18	9	0
16	100	88	38	6
17	100	85	58	13
18	100	59	90	0
19	67	85	0	0
20	100	33	67	44
21	96	100	8	0
22	100	45	58	52
23	86	92	40	28
24	85	100	92	14
Mean	93	69	46	14

Each group contained five mice; each mouse in one group

received a dose of Ext. Filicis B.P. which was 10 mg. for the first group, 5 mg. for the second group, 2·5 mg. for the third group, and 1·25 mg. for the fourth group.

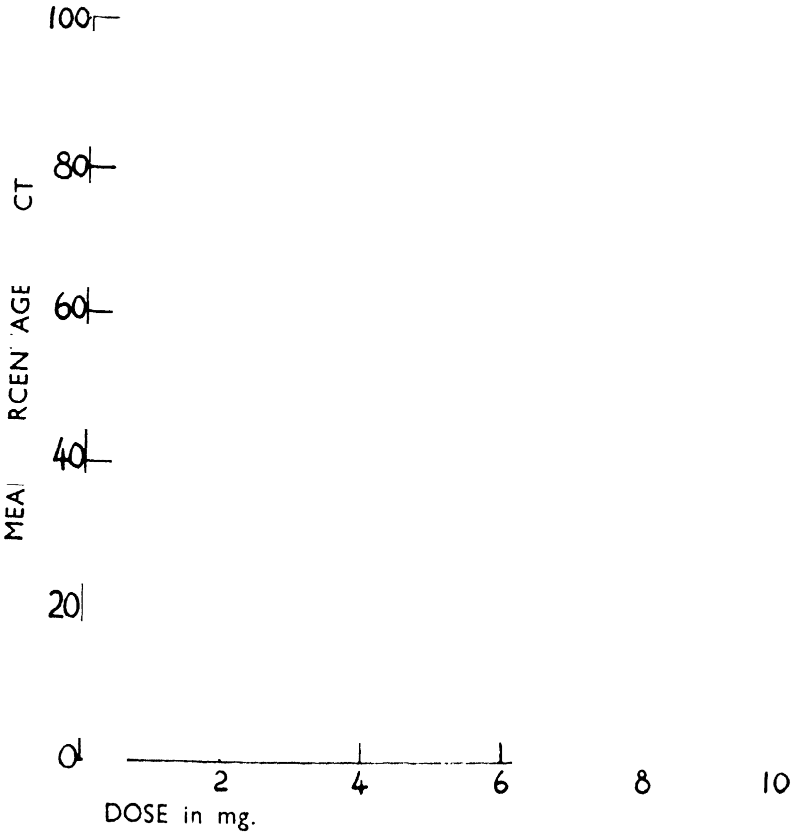


FIG. 70. Results of 11 experiments using a total of 55 mice for each dose of Ext. Filicis. Ordinates: mean percentage wet weight of tapeworms affected. Abscissae: dose (mg.) given to each mouse. (Holton, *Brit. J. Pharmacol.* 1947, **2**, 100.)

The effect of the anthelminthic is then expressed as the percentage of affected worms. For example, in Table LXXIV the mice of Group 2 contained a total of 481 mg. worms, these being worms in the small intestine, the large intestine, and in the faeces. Of this total, all were affected by the drug except the living worms in the small intestine which weighed 75 mg. Thus

the affected worms weighed 406 mg. and the percentage of affected worms was $\frac{406}{481} \times 100$, or 85.

When she repeated experiments of this kind using the same

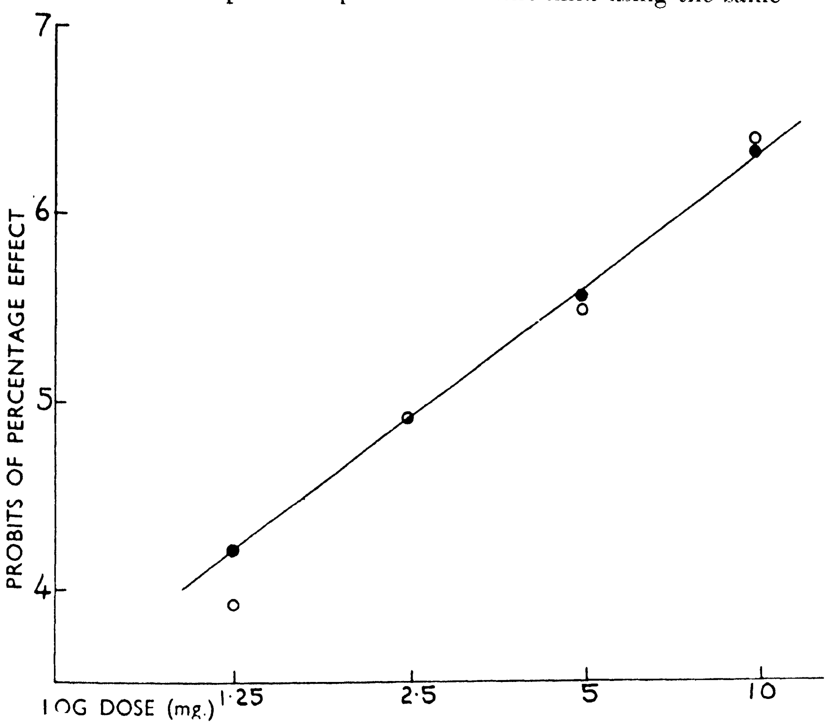


FIG. 71. Same results as in Fig. 70. Ordinates: probits of mean percentage wet weight of tapeworms affected. Abscissae: log dose (mg.) given to each mouse. White circles: empirical probits. Black circles: expected probits. (Holton, *Brit. J. Pharmacol*, 1947, **2**, 100.)

doses of the same Ext. Filicis in each, Holton obtained the results shown in Table LXXV.

In any single experiment of Table LXXV the response is not necessarily graded to the dose, but if the results of any three consecutive experiments are taken together the mean percentage wet weights of tapeworms affected are graded to the doses. Therefore, at least fifteen mice should be used for each dose of an anticestode drug. The result is shown graphically in Figs. 70 and 71. In Fig. 71 the logarithm of the dose is plotted against the probit of the mean percentage effect.

received a dose of Ext. Filicis B.P. which was 10 mg. for the first group, 5 mg. for the second group, 2·5 mg. for the third group, and 1·25 mg. for the fourth group.

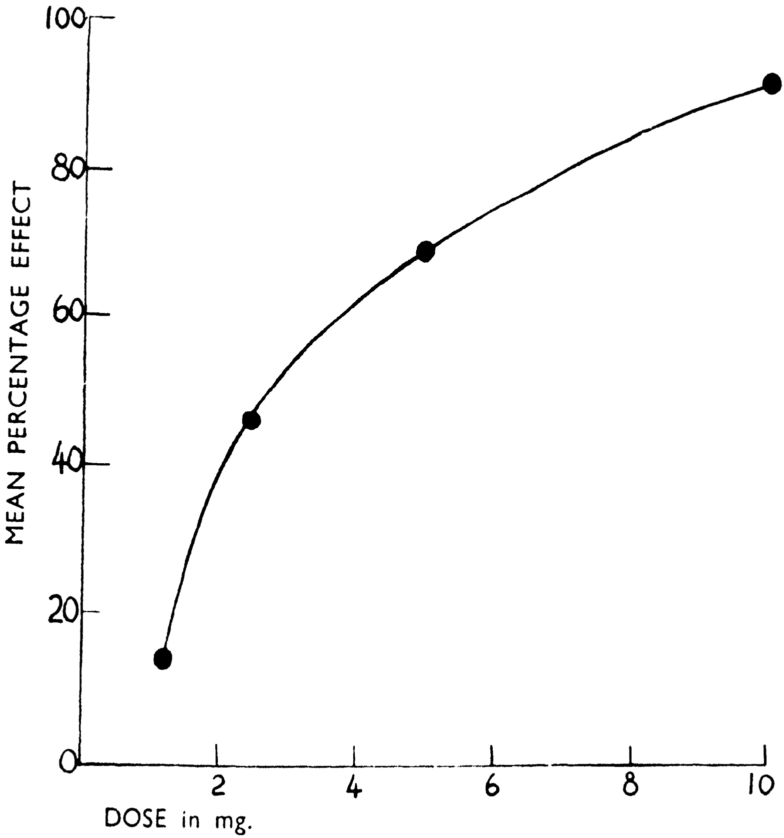


FIG. 70. Results of 11 experiments using a total of 55 mice for each dose of Ext. Filicis. Ordinates: mean percentage wet weight of tapeworms affected. Abscissae: dose (mg.) given to each mouse. (Holton, *Brit. J. Pharmacol.* 1947, **2**, 100.)

The effect of the anthelminthic is then expressed as the percentage of affected worms. For example, in Table LXXIV the mice of Group 2 contained a total of 481 mg. worms, these being worms in the small intestine, the large intestine, and in the faeces. Of this total, all were affected by the drug except the living worms in the small intestine which weighed 75 mg. Thus

the affected worms weighed 406 mg. and the percentage of affected worms was $\frac{406}{481} \times 100$, or 85.

When she repeated experiments of this kind using the same

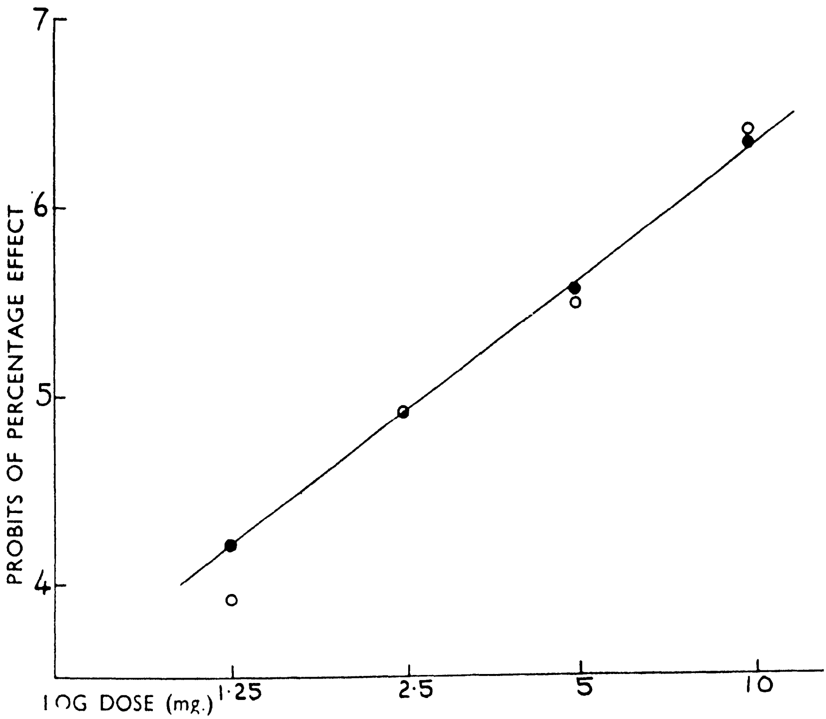


FIG. 71. Same results as in Fig. 70. Ordinates: probits of mean percentage wet weight of tapeworms affected. Abscissae: log dose (mg.) given to each mouse. White circles: empirical probits. Black circles: expected probits. (Holton, *Brit. J. Pharmacol.* 1947, **2**, 100.)

doses of the same Ext. Filicis in each, Holton obtained the results shown in Table LXXV.

In any single experiment of Table LXXV the response is not necessarily graded to the dose, but if the results of any three consecutive experiments are taken together the mean percentage wet weights of tapeworms affected are graded to the doses. Therefore, at least fifteen mice should be used for each dose of an anticestode drug. The result is shown graphically in Figs. 70 and 71. In Fig. 71 the logarithm of the dose is plotted against the probit of the mean percentage effect.

REFERENCES

Bertram and Gordon, 1939, *Ann. Trop. Med. Parasitol.* **33**, 279.
Culbertson and Rose, 1944, *J. Pharmacol.* **81**, 189.
Hawking and Sewell, 1948, *Brit. J. Pharmacol.* **3**, 285.
Holton, 1947, ibid. **2**, 100.
Hunt and Davey, 1947, *J. Trop. Med. Hyg.* **50**, 53.
Schubert, 1948, *American J. Trop. Med.* **28**, 121, 137, 157.
—— 1949, ibid. **29**, 115.
Standen, 1949, *Ann. Trop. Med. Parasitol.* **43**, 13.
Williams and Brown, 1945, *Science*, **102**, 482.

CHAPTER XXII

TRYPANOCIDAL SUBSTANCES

EXPERIMENTAL trypanosome infections may be used for two distinct purposes:

(*a*) For the discovery and assay of new drugs designed to be used for the treatment of corresponding trypanosome infections in man or domestic animals.

(*b*) For the routine control of potency of drugs which cannot be standardized satisfactorily by chemical or physical methods. In such cases the test is a means of ensuring the uniformity of different samples of drug and need bear no relation to the use to which the substance will be put in medicine. The control tests for Neoarsphenamine and Sulpharsphenamine are of this type and will be discussed fully on p. 390.

GENERAL PRINCIPLES

The principle of the test for trypanocidal activity is as follows: Mice are inoculated with a suspension of blood containing trypanosomes. They are treated with drug, which may be given before or immediately after inoculation, or after a period during which the infection is allowed to become established in the mouse. The effect of the treatment is assessed by observing either the disappearance of trypanosomes from the peripheral blood or the survival times of the infected mice.

Trypanosome infections in mice. The characteristics of some trypanosome infections which may be used for the infection of mice by blood inoculation are listed in Table LXXVI.

Maintenance of strain. When a strain of trypanosome is isolated from its natural host and transferred to laboratory animals for the first time it is unusual for all the animals inoculated to become infected. Several passages in the mouse, rat, or guinea-pig are often required before the strain becomes adapted to the new host and attains sufficient virulence to kill. When once established in mice the strain may be maintained indefinitely by passage. The pathogenicity may often be increased by frequent transference using heavy inocula by the

TABLE LXXVI

Trypanosome	*Natural host*	*Disease produced*	*Characteristics of mouse infection*
T. equiperdum .	Horse, donkey.	Dourine.	Intense infection of peripheral blood, which is lethal in 2–4 days.
T. brucei . .	Cattle and other domestic animals.	Nagana.	,,
T. rhodesiense . *T. gambiense* .	Man.	Sleeping-sickness.	,,
T. evansi .	Horse, camel, cattle.	Surra.	,,
T. equinum .	Horse, mule, ass.	Mal de Caderas.	,,
T. congolense .	Cattle, goat, sheep, pig.	. .	Some strains give a heavy infection which kills in 3–10 days; others give chronic infections in all or some of the mice, and trypanosomes may disappear from the peripheral blood spontaneously.
T. cruzi . .	Man, dog, cat.	Chagas' disease.	Chronic infection which takes 5–14 days to appear in the peripheral blood; varies from day to day and may disappear for several days. Sometimes kills in 20–30 days.

peritoneal route. With *T. congolense* the characteristics of a strain may be permanently changed by allowing the infection to become chronic before passage into a fresh mouse. Browning and Calver (1943) have shown that in order to keep the properties of *T. congolense* constant it is essential that passage should be performed at the fastigial stage, that is, when the trypanosomes are multiplying rapidly in the blood for the first time and before there has been time for an immunity response to develop.

The laboratory strain of *T. equiperdum*, which kills mice within a day or two, is often kept in guinea-pigs as a standby. In this host the infection is a chronic one, and the animal maintains a low-grade peripheral blood infection for several weeks. A few passages in mice will quickly restore the virulence,

because this organism stimulates only slight immunity response.

A convenient method of passage, which does not involve the killing of the infected mouse, is to take a tiny drop of blood straight from the tail into the broken end of a stout but finely drawn Pasteur pipette. The blood may then be inoculated with the pipette directly under the skin or into the abdominal cavity of a 'clean' mouse. The subcutaneous route gives a rather less rapidly developing infection than the intraperitoneal.

In a laboratory where a number of different strains are kept it is very important that a strict routine and a colour-marking system should be used so that there is no chance of the strains becoming confused. A spot of carbolfuchsin or alcoholic picric acid applied to the head or back of the mouse is a satisfactory method of marking.

Kolmer's text-book (1926) contains a great deal of information upon methods of inoculation and other matters concerning trypanosome tests, and is well worth consulting.

Great care should be taken when dealing with the human pathogens, particularly *T. cruzi*, which can penetrate the mucous membranes and conjunctiva. If an accident occurs in the laboratory, and there is a possibility of a worker having been inoculated with *T. rhodesiense* or *T. gambiense*, a dose of suramin (B.P.) 1 gm. intravenously, or of pentamidine 0·15 gm. intramuscularly, should be given.

Inoculation of animals for test. In some laboratories it is customary to use an infected rat for the preparation of the inoculum, because more blood may be obtained from the larger animal. This is not generally necessary, however, because from a single mouse heavily infected with strains such as *T. equiperdum* or *T. rhodesiense*, sufficient blood may be obtained to inoculate 200–300 mice, each with a dose of about 2 million organisms. If more are required, two or three mice may be used. The use of the mouse for the inoculum also avoids a change of host, though the pathogenicity is but little affected by passage in the rat.

The infected animal is anaesthetized and bled from the heart. The blood is diluted to contain a standard number of trypanosomes, and injected into the groups of test mice without delay.

The most widely used diluent is citrated saline (0·9 per cent. sodium chloride, 1·0 per cent. sodium citrate), to which glucose (1·0 per cent.) is sometimes added. This is not ideal because sodium citrate has a toxic action upon trypanosomes, and citrated saline renders *T. rhodesiense* immobile in less than 2 hours, and non-infective for mice in about 12 hours. Infectivity is reduced after standing for only 4 hours. This is not a matter of great moment when mice are being inoculated within a short time of preparation of the suspension, but it is a good principle to do as little harm to the organisms as possible during transference. Heparin is to be preferred to citrate as an anticoagulant, because it is much less toxic to the trypanosomes. When the dilution is considerable (1 in 30 or more), it is not necessary to use any anticoagulant at all; the small web of fibrin which separates does not interfere if the suspension is well stirred. Trypanosomes survive for 24–48 hours or longer at room temperature when suspended in plain saline. At 37° C. it is necessary to have rabbit or horse serum (10–15 per cent.) present in the suspension in order for the trypanosomes to survive for more than an hour or two. Some workers take the precaution of adding horse or rabbit serum to all suspensions of trypanosomes. Human serum is toxic to some strains and should not be used.

Administration of drugs. Solutions or suspensions of drugs may be given by subcutaneous, intraperitoneal, intramuscular, or intravenous injection, or by mouth. For the detection of traces of activity, the first dose of drug may be given before the trypanosomes are injected, and further doses given once or twice daily until the mouse dies from the infection or a protective effect is evident (Fig. 72). In such a test it is undesirable to inject both drug and trypanosomes intraperitoneally as the test then assumes the character of an *in vitro* experiment, the peritoneal cavity of the mouse taking the place of the test-tube to which drug and organisms are added.

If the substance being assayed is known to have activity, a single dose is usually all that is required to produce a measurable effect. In this case there are two general methods to choose from. Gray, Trevan, Bainbridge, and Attwood (1931) and Bülbring and Burn (1938) described tests in which the dose of drug was given shortly after the inoculation of trypanosomes,

the infectivity of the inoculum being checked by means of a group of untreated, infected control mice. These methods are therefore prophylactic in nature. Voegtlin and Smith (1920), Hawking (1941), and others have used curative methods in which the animals, after inoculation, are kept for 1 or 2 days before the injection of the drug, and the presence of trypano-

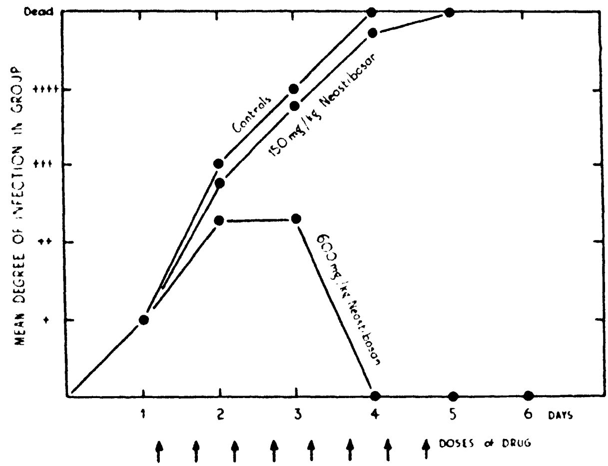


FIG. 72. Groups of mice infected with *T. equiperdum* were given subcutaneous doses of neostibosan twice a day. 150 mg./kg. had practically no effect; 600 mg./kg. cleared the infection from the peripheral blood after 3 days treatment.

somes in the peripheral blood of each assured before giving the dose. Voegtlin and Smith (using rats) went to the length of counting the trypanosomes in the blood of each animal, and selecting those containing 100,000–300,000 per μl. for the test. It is generally unnecessary to go to the enormous labour that counting entails, and it is sufficient for most practical purposes to examine a drop of blood directly under a coverslip.

For preliminary tests on new drugs, groups of five mice are generally large enough to give a reliable indication of presence or absence of activity. It is essential that every test should be controlled with a group of infected untreated controls; tests in which these are omitted are worthless. For official potency tests (see p. 390), and for assays of a drug in comparison with a standard, groups of at least ten mice must be used, and several dose-levels of both standard and test are necessary if calculation

of the potency is to be made. The dose-levels should be in *geometric series.*

Assessment of trypanocidal effect. There are two criteria which are commonly used for the assessment of trypanocidal activity and the calculation of the relative potency of drugs:

1. The proportion of mice in which trypanosomes disappear from the peripheral blood as a result of the dose of drug;
2. Estimation of the survival time of the group of mice treated.

1. *Disappearance of trypanosomes from the peripheral blood.* Daily examination of blood from the tails of animals following an effective dose of trypanocide shows that trypanosomes disappear after an interval which varies with the nature of the substance given. Thus, with tervalent antimonials such as tartar emetic, or stibophen, all organisms have disappeared in 3 or 4 hours; whereas with suramin or phenanthridinium compounds, the trypanosomes may at first increase in numbers and disappear only after 2 or 3 days (Fig. 73).

By trial and error with different doses of drug the smallest dose which 'clears' all the mice can be found, and this has been called the 'minimum effective dose'. At such a dose-level all of the mice subsequently relapse. A larger dose, the 'minimum curative dose', may be determined, which maintains the peripheral blood free from organisms for a long period, usually taken as 28 days. These figures, based upon the original work of Ehrlich, have been used extensively for the description of trypanocidal activity in the very fruitful researches of Warrington Yorke and his school, and of Carl Browning. Much criticism has been levelled at such figures, because they vary from test to test and because the error cannot be calculated statistically. Nevertheless, a record of experimental results similar to that compiled in Table LXXVII does indicate clearly the approximate range over which a drug will act, especially if a standard drug is used in the same test for comparison. For drugs which show some activity, but are low in potency compared with a standard of the same chemical type, such a description of the range of activity will often suffice. When it is necessary to compare two substances of great potency, and to be sure of the relationship between their activities and of the significance of the result, a more elaborate method must be applied.

Hawking has applied the process of probit analysis to the observed proportions of mice 'cleared' of trypanosomes in the peripheral blood by a range of doses increasing by logarithmic intervals. This method has been used for the description of the action of suramin (1941) and for the standardization of potency of the official arsphenamines (1943). The test laid down in the

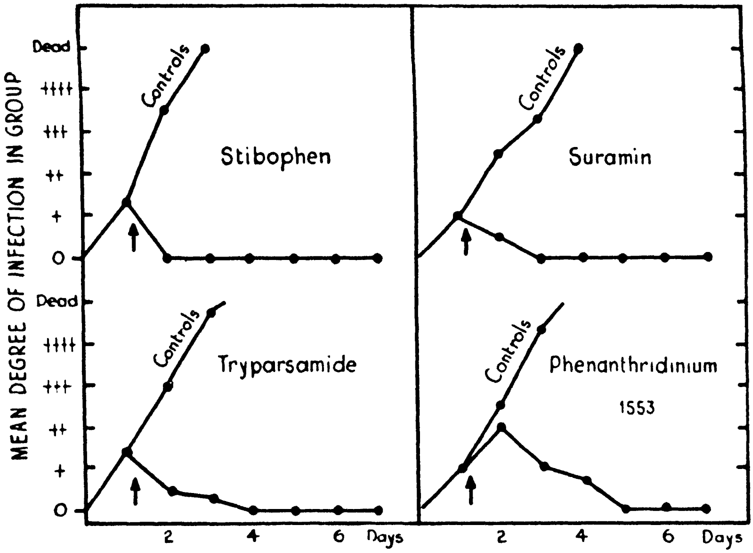


FIG. 73. The different rates at which infections of *T. equiperdum* in groups of mice respond to curative doses of different types of trypanocide. Single doses were given subcutaneously on the day following inoculation; stibophen 100 mg./kg., suramin 1·5 mg./kg., tryparsamide 400 mg./kg., phenanthridinium (1553) 50 mg./kg.

British Pharmacopoeia, 1948, which is based upon this work, is described on p. 391.

The probit of the percentage of animals 'cleared' after a suitable time interval of 1–4 days is plotted against the logarithm of the dose. The median effective dose (the dose which would be expected to 'clear' 50 per cent. of the animals) is estimated graphically, or by calculation, and its error is determined. The potency of a preparation compared with that of a standard can be calculated, and the fiducial limits of the ratio obtained can be determined by the method described in Chapter III, Section IX. The test as it stands measures only the immediate activity of the drug in causing trypanosomes to

disappear from the blood, and the dose range used is far below that needed to produce permanent cure. For the determination of the *curative power* of a drug the calculation may be applied to the proportion of animals which remain free from trypanosomes for several weeks. For example, Chen, Geiling, and MacHatton (1945) compared the curative activities of antimony compounds by recording the percentages of mice which survived for 21 days after treatment. The probit of the percentage cured was plotted against the logarithm of the dose and the median curative dose determined from the curve.

2. *Survival time.* Gray, Trevan, Bainbridge, and Attwood (1931) describe a method in which the effects of doses of organic antimonials are measured by the mean survival times of the groups of mice treated. This test has the advantage of requiring little attention while it is in progress beyond counting the number of animals which have died from relapse of the infection from day to day. It has the disadvantage of taking longer to complete, and the possible source of trouble from intercurrent infections in the mice. The latter point may be checked by examination of the blood of dying or freshly dead mice, to ensure that trypanosomes are present.

A development of this method was described by Bülbring and Burn (1938). In their experiments the dose-levels were chosen so that all mice treated with the lowest dose were dead shortly after the group of infected, untreated controls. The number of mice surviving each day was counted, up to a maximum of 6 or 8 days, and the daily figures totalled to give a number representing the degree of protection afforded by the dose during the test period. A simple calculation will show that this total, in the case of groups which succumb before the end of the test, is a measure of the survival time of the group, multiplied by the number of mice used. The relative activities of drugs may be assessed by plotting the survivor-totals against the logarithms of the doses. This method of treatment has the great advantages of speed and simplicity. It gains in accuracy if the survivors are counted at 8- or 12-hourly intervals.

The Bülbring and Burn totals may be used to calculate the relative potencies of two drugs by the method described in Chapter III, Section VI.

Alternatively, drugs may be compared by the determination

of the dose of each required to produce a standard effect. Goodwin (1944) has taken the 'standard effective dose' as the dose which gives a survival time midway between that of untreated controls and the arbitrary end of the test. Chen, Geiling, and MacHatton (1945) have chosen the 'standard suppressive dose' as the dose which prolongs the life of treated mice by a time equivalent to the survival time of the untreated controls. Both of these figures are readily determined from the dose-response curves.

With antimony compounds estimates of potency by 'clearance', 'cure', and survival time give very similar results, so long as the substances compared are of the same chemical type.

Experimental details for tests with *T. equiperdum*, *T. brucei*, *T. gambiense*, *T. rhodesiense*, *T. evansi*, or *T. equinum*. A heavily infected mouse is placed in a jar with a pad of cotton-wool and some ether. When the respirations have just ceased the animal is removed, pinned on a board, the fur swabbed with antiseptic, and the heart well exposed by cutting away the sternum and the adjoining portions of the ribs. A sterile teat pipette which has been drawn out and broken to a short fine point is introduced through the wall of the right ventricle, and blood is withdrawn slowly, while the heart is still beating. About 1·5 ml. of blood may be obtained in this way from a mouse weighing 20 gm. The blood is discharged, before it clots, into a vessel containing about 50 ml. of sterile normal saline, and stirred so that the fibrin separates. A drop of the stirred suspension is examined in a haemocytometer cell and the number of trypanosomes per μl. determined. It is usually unnecessary to fix and stain the organisms because they become attached to the upper and lower glass surfaces and do not change their positions in the field to any great extent. The wriggling of the trypanosomes makes them easy to see.

For determining the number of trypanosomes per μl. in undiluted blood, or in a concentrated suspension containing relatively large numbers of red cells, dilution may be made in a haemocytometer pipette, with a fluid containing acetic acid and a dye similar to that used for the enumeration of leucocytes. The red cells are lysed and the trypanosomes stained. However, a more accurate figure is given when saline is used as the diluent,

because trypanosomes as well as red cells are lysed by the acetic acid and the result is always too low.

The primary suspension is diluted with saline so that the required concentration is obtained. An inoculum of 0·5 ml. containing 4,000 trypanosomes per μl. (2 million organisms) is satisfactory, but more may be given if it is found by experience that the infection is a little too slow in developing. The number inoculated is not very important, provided that sufficient trypanosomes are given to produce an infection which develops quickly enough to smooth out small differences in susceptibility of the individual mice.

The mice used for the test should be uniform in weight and strain. They are injected intraperitoneally with the suspension, care being taken that none of the inoculum escapes. It is best to withdraw food from the cages overnight and to inoculate first thing in the morning when the bellies of the mice are not distended.

The suspension should be stirred frequently, and a drop should be examined under the microscope when the injections have been completed, to confirm that the organisms are still motile.

(*a*) If the test is to be a screening test for new drugs, the first dose of drug is given forthwith. Doses are given once or twice daily, and should be large enough to approach the toxic range so that the drug is given every opportunity of showing any activity it may possess. It is usually safe to give repeated doses of one-half of the L.D. 10, calculated from a preliminary toxicity test. The subcutaneous route is the most convenient. If desired the first dose may be given before inoculation with trypanosomes. A group of five mice is large enough for a preliminary experiment; one group is kept untreated as a control.

The peripheral blood of each mouse is examined daily for trypanosomes. The tip of the tail is snipped with scissors and a small drop of blood, transferred to a slide and covered with a ½ in. × ½ in. coverslip. The size of the drop of blood taken from each mouse should be as nearly as possible the same.

The erythrocytes should be in a single layer but not widely separated, as in a good blood film. Under the ⅛th-inch objective, the approximate degree of infection may be assessed as follows:

1 trypanosome in about 10 fields . .	+ or '1'
1–5 trypanosomes per field . . .	++ or '2'
5–50 „ „ . . .	+++ or '3'
Trypanosomes about equal in numbers to the erythrocytes	++++ or '4'.

The infection in the controls mounts rapidly from about ++ on the day after inoculation to ++++ and death in 2–4 days, and by comparison it is easy to see which drugs have had an effect. Treatment with drug is continued until the controls are dead, and any substance with trypanocidal action is investigated more fully by giving single or repeated doses of graded size to larger groups of animals.

Fig. 72 shows the development of infection after treatment with repeated doses of neostibosan.

(*b*) If two samples are to be compared by the method of Bülbring and Burn, the animals in groups of ten or more are given doses of drug by intravenous, intraperitoneal, or subcutaneous injection as soon as the inoculations have been completed. At least two dose-levels should be used for each drug to be tested. The mice are set aside, and the numbers of survivors in each group counted every day until the end of the test in 6 or 8 days. The controls die in 2–4 days, and, if the doses have been chosen correctly, one group of the treated mice should survive almost to the end of the test, and the others for periods graded between this and the survival time of the controls. The daily numbers of survivors are totalled as in Table LXXVIII and the totals when plotted against the logarithms of the doses, give curves which may be used to estimate the relative potencies of the drugs.

(*c*) If Hawking's 'clearance' method is to be used, the inoculated animals are fed and set aside for 24 hours. The blood of each is then examined, and any mouse showing an unusually light or heavy infection is rejected. If the examination of a few mice shows that the infections are all of low intensity, it is best to set the animals aside for a further 24 hours.

The mice are divided into groups of ten or twenty, and one group kept as a control. The other groups are given doses of the drugs to be compared. At least two dose-levels for each drug should be given. The blood of each is re-examined 1, 2, or 3 days later, depending upon the time taken for the drug to exert its maximum effect.

The percentage of mice in which the peripheral blood is cleared of trypanosomes is determined at each dose-level. When searching for trypanosomes it is sufficient to examine carefully twenty fields under the $\frac{1}{8}$th-inch objective. If no trypanosomes are seen, the mouse is recorded as 'cleared'.

The probit of the percentage of mice cleared is plotted against

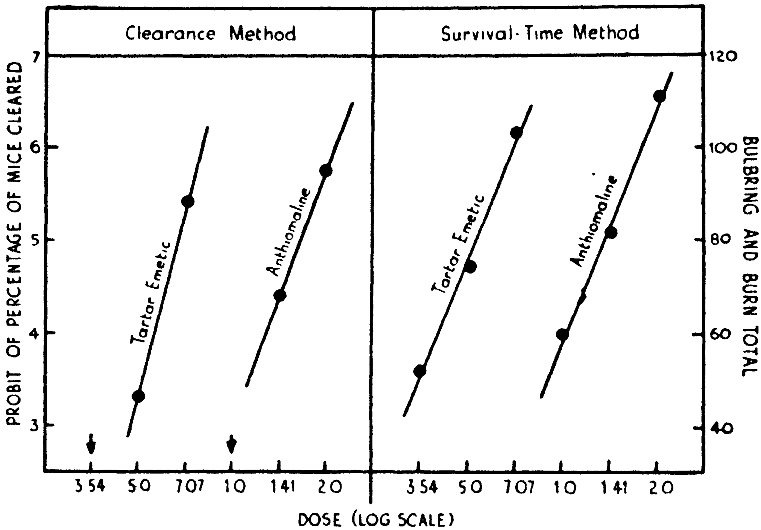


FIG. 74. The results of the test recorded in Tables LXXVII and LXXVIII are here shown plotted side by side. The potency ratio of the two drugs is the same by both methods of assessment.

the logarithm of the dose, and the relative potencies of the drugs determined graphically or by calculation.

It is desirable to obtain as much information as possible from every group of animals used, and it is very little extra trouble to keep the groups of mice for a further week and to determine the survival times, by the method of Bülbring and Burn. Sometimes an unhappy choice of dose-levels yields maximal or minimal responses which cannot be used for calculation by the clearance method, but the test may yet yield useful information if the mice are kept for determination of survival times.

In the following example both the clearance and survival-time methods have been applied to a comparison of the activities of Tartar emetic and Anthiomaline against *T. equiperdum* (see Fig. 74).

Example. Comparison of the activity of 'Anthiomaline' and Tartar emetic. Mice were inoculated intraperitoneally with 0·5 ml. of a suspension of *T. equiperdum* containing 4,000/μl. The inoculations were completed in 30 minutes. Infection was confirmed by blood examination on the following day, and doses of drugs given subcutaneously.

TABLE LXXVII. *Numbers of mice cured, relapsing, and unaffected*

	Tartar emetic			*Anthiomaline*			*Controls*
Dose mg. kg.	*3·54*	*5·0*	*7·07*	*10·0*	*14·14*	*20·0*	..
Total no. of mice	19	19	20	20	21	18	13
Permanently cured	0	0	2	0	0	1	0
Cleared, but relapsed	0	1	11	0	6	13	0
Not cleared	19	18	7	20	15	4	13

TABLE LXXVIII. *Clearance and survival-time figures*

Drug	*Dose, mg. kg.*	*Number of mice*	*Number cleared*	*Daily survivors* *1*	*2*	*3*	*4*	*5*	*6*	*Bulbring and Burn totals (corrected to apply to 20 mice)*
Tartar emetic	3·54	19	0	19	19	11	0	..	..	52
	5·0	19	1	19	19	19	6	4	3	74
	7·07	20	13	20	20	20	15	14	14	103
Anthiomaline	10·0	20	0	20	20	16	3	1	0	60
	14·14	21	6	20	20	18	10	9	8	81
	20·0	18	14	18	18	18	18	15	13	111
Controls	..	13	0	13	13	3	0	..	..	45

Results of calculations.

(*a*) *From the percentage cleared* (see Section IX, Chapter III):
1 mg. Anthiomaline is equivalent to 0·407 mg. Tartar emetic.
Limits ($P = 0·95$): 87 to 115 per cent.

(*b*) *From the Bulbring and Burn totals* (see § 59, Section VI, Chapter III):
1 mg. Anthiomaline is equivalent to 0·393 mg. Tartar emetic.
Limits ($P = 0·95$): 89 to 112 per cent.

Special technique for *T. congolense* infections. With *T. congolense* infections, trypanosomes never appear in the peripheral blood in such large numbers as with *T. equiperdum*. The inoculum used is therefore less heavy and 1,000–2,000 per μl. is a fair average. Larger numbers may be obtained from a rat or by pooling the blood from several mice. Some strains are sufficiently virulent to kill all infected mice in 5–10 days, and

then both 'clearance' and survival-time methods may be applied. About 21 days is often necessary to obtain satisfactory survival-time figures, and frequent blood examinations must be made to avoid errors resulting from intercurrent infection. Other strains of *T. congolense* never kill mice. The clearance method may be used, or, alternatively, the number of days during which the blood remains free from trypanosomes may be determined and used as a graded response. This, of course, involves the examination of the blood each day. It is a tedious business, but essential if the most is to be made of a test using *T. congolense*. Some very chronic strains are liable to disappear spontaneously from the peripheral blood for days at a time, and then great care must be taken to compare the results of drug action with the untreated controls at every point. Browning and Calver (1943) and Calver (1945) have discussed at length the factors affecting the behaviour of *T. congolense* in the mouse. Wien (1946) has made careful comparisons of the action of phenanthridinium compounds against strains of *T. congolense* of different degrees of chronicity. Table LXXIX

TABLE LXXIX. *The effect of two phenanthridinium compounds upon the survival times of mice infected with T. congolense*

Drug	*Dose, mg. kg.*	*Survivors (days after inoculation)*														*Bulbring and Burn totals*
		3	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	
Phenanthridinium 897	0·4	10	10	8	8	8	8	7	7	7	7	6	5	5	4	100
	0·1	10	10	5	5	4	4	3	3	1	1	1	1	1	1	50
	0·025	10	10	4	4	1	1	1	1	1	1	1	1	1	1	38
Phenanthridinium 1553	0·4	10	10	10	10	10	10	10	9	9	9	9	8	8	6	128
	0·1	10	10	8	8	8	8	7	7	7	6	3	3	3	2	90
	0·025	10	10	5	2	2	2	2	2	2	1	1	0	..	..	39
Controls	..	10	10	3	1	1	1	1	0	..	..	..	..	..	..	27

A single dose was given on the third day after inoculation, when trypanosomes were present in the blood.

shows a typical set of results obtained from a test on phenanthridinium compounds.

Special technique for *T. cruzi* infections. *T. cruzi* produces a chronic infection in mice. Trypanosomes do not appear in the peripheral blood for about 5–12 days and never become very numerous. They may disappear for considerable periods. Mice sometimes die in 14–28 days, but may live for many weeks and develop a peculiar gelatinous oedema (Collier, Fulton, and Innes, 1942). The life-cycle of the parasite is different from

those of the species of trypanosomes described above, because the flagellates enter host cells in the skeletal or cardiac muscle and multiply there as leishmanial forms. No division takes place in the peripheral blood. Infection is most readily produced by the subcutaneous injection of a small volume of a concentrated inoculum. An inoculum of 0·1 ml. containing about 500 organisms per μl. is satisfactory. Heavy infections are produced in new-born rats, but it is seldom convenient to make use of these animals for routine tests. An effective drug must kill the intracellular forms of the parasite, and so far the only drugs found to have any action at all are the quinoline derivative 'Bayer 7602' ('Chagavlon') and some phenanthridinium compounds. Slight activity of drugs may best be detected by giving a series of daily doses during the latent period of 5–12 days immediately after inoculation. Table LXXX shows a typical experiment using 'Bayer 7602'.

TABLE LXXX

		Degree of infection on successive days after inoculation													
Drug	*Mouse*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*
Bayer 7602, 150 mg. kg. for 6 days	1	+	:	+	+	+	0	0	0	0	0	0	0	0	0
	2	+	+	+	+	+	0	0	0	0	0	Dead	—	—	—
	3	+	+	+	+	+	+	:	+	:	+	+	+	+	+
	4	:	+	+		+	+	0	0	0	0	0	0	0	0
	5	+	0	:	+	+	0	:	0	0	0	0	0	0	0
	6		0	+	+	0	0	0	0	0	0	0	0	0	0
	7	+	0	+	+ +	+	+	0	:	0	0	0	0	0	0
	8	+	0	+	+	+	0	0	0	0	0	0	0	0	0
Controls	1		:	+	+ +	+ +	+ + +	Dead	—	—	—	—	—	—	—
	2	:	+	+	+ +	+ :	+ +	+	+ + +	+ + +	+ + +	Dead	—	—	
	3	+	0	0	0	+	0	0	0	0	0	0	0	0	Dead
	4		+	+	0	+	+	+	Dead	—	—		—		—
	5	+	+	+	+	+	+	+ +	Dead	—	—	—	—	—	—
	6	:	0	0	Dead	—	—	—	—	—	—	—	—		—
	7	:	:	:	+	:	+ +	:	+	+ +	+	+ +	Dead	—	—

The degree of infection with *T. cruzi* in the blood of individual mice. One group was treated with Bayer 7602, given subcutaneously for 6 days beginning on the eighth day after inoculation.

' : ' indicates that a very few organisms were detected in the blood after prolonged search.

In vitro METHODS

It is quite simple to test the trypanocidal activity of a drug which kills trypanosomes quickly, by preparing suspensions of organisms in saline or Ringer's fluid containing serial dilutions of the drug, and keeping at room temperature. The rate at which the organisms die, or the limiting concentration in which

they will survive for a few hours, may be used as a measure of the potency of the drug.

This method is quite valueless for substances which need to be metabolized in the mammalian host before they become active, and is also useless for drugs which take time to act. Yorke, Adams, and Murgatroyd (1929) devised a method of keeping trypanosomes alive for 24 hours *in vitro* at 37° C. and were enabled thereby to demonstrate the direct trypanocidal action of the drug synthalin, the forerunner of the aromatic diamidines.

The best method to use, is first to inactivate some rabbit, sheep, or horse serum by heating for 30 minutes at 62°–63° C. It is a good thing to centrifuge the serum after inactivation, to remove any particulate matter. Kahn tubes are used for the test, and 0·5 ml. of serum is placed in the first tube with 0·5 ml. of a solution of the drug to be tested in Ringer-glucose (sodium chloride 0·9 per cent.; potassium chloride 0·025 per cent.; calcium chloride 0·02 per cent.; sodium bicarbonate 0·015 per cent.; glucose 0·5 per cent.). Quantities of 0·5 ml. of a mixture of equal parts of serum and Ringer-glucose are distributed into the other tubes, and serial dilutions of the drug prepared in the usual way, from the highest concentration in the first tube. Some tubes are kept without drug as controls. The blood of a mouse infected with *T. rhodesiense* is taken aseptically, and diluted with an equal volume of citrated saline in a Kahn tube. The blood is centrifuged, and the white creamy layer of trypanosomes pipetted from the surface of the packed erythrocytes. The trypanosomes are resuspended in serum and Ringer-glucose medium, and are again centrifuged. The washing is repeated and the organisms are finally suspended and diluted so that 0·025 ml. contains about 500,000 trypanosomes. To each tube of medium and drug is added 0·025 ml. of the suspension, and the tubes are incubated at 37° C. All of these operations must be carried out aseptically. Yorke, Adams, and Murgatroyd recommended that the tubes be covered with small glass caps because cotton plugs were thought to affect the trypanosomes adversely.

The number of living trypanosomes present in each tube is determined by taking a drop of the stirred suspension and counting in a haemocytometer chamber. This is done at the

beginning of the experiment, and at intervals during the following 24 hours. Table LXXXI shows how trypanosomes are killed by organic arsenic compounds when tested by this method.

TABLE LXXXI. *The effect of Tryparsamide and* p-*aminophenylarsenoxide upon* Trypanosoma rhodesiense in vitro (*Yorke & Murgatroyd* (1930))

Drug	*Concentration of drug in medium*	*No. of living trypanosomes per 256 squares of the haemocytometer scale*							
		Start	*2 hrs.*	*4 hrs.*	*8 hrs.*	*12 hrs.*	*16 hrs.*	*20 hrs.*	*24 hrs.*
Tryparsamide	1:100		46	24	0	..	..	..	..
	1:200		..	62	0	..	..	..	..
	1:400		..	..	37	0	..	..	..
	1:800		..	..	66	53	52	0	..
	1:1,600		..	..	..	81	51	0	..
	1:3,200		..	..	..	..	71	68	74
	1:6,400		..	..	..	..	..	..	90
p-aminophenyl-arsenoxide	1:12,800,000	66	0	..	..	..	..	..	..
	1:25,600,000		6	0	..	..	..	..	..
	1:51,200,000		32	0	..	..	..	..	..
	1:102,400,000		52	36	0	..	..	..	..
	1:204,800,000		..	52	21	0	..	..	..
	1:409,600,000		..	..	78	1	0	..	..
	1:819,200,000		..	..	..	104	84	76	82
	1:1,638,400,000		..	..	..	..	86	90	90
	Control		..	..	..	..	92	96	90

The results obtained by the method described above are apt to be erratic. Some samples of serum support the trypanosomes better than others, and changes in pH which sometimes occur during the test have a profound effect. If the sample of serum is found to be very alkaline, it is a good thing to adjust the pH to 7·4–7·8 by saturation of the medium with carbon dioxide before addition of the trypanosomes.

DETAILS OF OFFICIAL TESTS FOR TRYPANOCIDAL SUBSTANCES

1. ARSPHENAMINES

The official arsphenamines, neoarsphenamine, and sulpharsphenamine are required by the *British Pharmacopoeia* to comply with tests for absence of undue toxicity and for trypanocidal potency. These tests are necessary because the drugs are not chemical identities and cannot be controlled by the chemical and physical measurements which are at present considered

adequate for the other official arsenicals, oxophenarsine and tryparsamide. Each batch of arsphenamine manufactured is therefore compared with the standard preparation obtainable from the National Institute for Medical Research, Hampstead. In the appendix of the *Pharmacopoeia*, 'suggested methods of test' are set forth in cumbrous legal phraseology. It might be questioned whether such wording is appropriate to a 'suggestion', but since every batch of drug before being issued for sale in Great Britain must also comply with the requirements of the Therapeutic Substances Act, and since this involves the submission of the results of tests to the National Institute for Medical Research before a licence can be issued, the 'suggested methods' amount in effect to legal requirements. Rats are used instead of mice in U.S.A. and Canada, and samples which pass the tests of the *British Pharmacopoeia* will not necessarily be acceptable in these countries.

(*a*) *Toxicity test*

The test is designed to ensure that samples with toxicity greater than 120 per cent. of the standard preparation do not pass the test. The researches of Durham, Gaddum, and Marchal (1929) laid the foundation for the method. In the *Pharmacopoeia* of 1932, the assumption was made that the mice used in any one laboratory would be uniform in their response to the standard preparation from time to time, and it was not therefore obligatory to compare the unknown with the standard at every test. The wisdom of this assumption was questioned by Burn in the previous edition of this book. In the 1948 *Pharmacopoeia* the test has been revised, and it is now necessary to inject one group of mice with the unknown and one with the standard for direct comparison in the same test. The test is simple and does not involve statistical calculation. It fulfils its purpose very well. The new method was worked out by Hawking (1943) and has actually been in use since 1941.

In order to conserve the supply of the standard preparation, it is usual for every laboratory performing routine tests to set aside a large batch of drug approximately equivalent in toxicity and potency to the standard preparation, and to compare new samples with this. This procedure is permissible provided that the laboratory standard has been approved by the Director of

Biological Standards at Hampstead as having been accurately assayed in terms of the standard preparation.

In the following description of the tests, reference to the 'standard' implies the standard preparation or its approved equivalent.

Method of the *British Pharmacopoeia*, 1948, for neoarsphenamine. Two groups of 20 mice are used; the fasting weights of individual mice must not vary by more than 3 gm.

The ampoule containing the standard in an atmosphere of nitrogen is opened, the drug is weighed rapidly and dissolved in freshly distilled water. A suitable concentration is 2·0 per cent., and the dose of the standard injected is chosen from experience to be sufficient to kill about 50 per cent. of the mice (usually 0·5 ml. of 2 per cent. solution per 20-gm. mouse). The dose given to every mouse is the same, and is adjusted to the average weight of the mice. Injections are given into the tail vein; it is essential that the whole of the dose should enter the vein and that the rate of injection should be approximately the same for each mouse. The mouse may be held by an assistant who stretches the tail towards the operator, or it may be held in a plugged cylinder of perforated zinc with the tail protruding from a notch in the cork. If the laboratory temperature is low, the tails may be immersed in water at 40° C. for a few moments to dilate the veins. The injection of the group of 20 mice must be completed in less than 10 minutes from the time of preparation of the solution of neoarsphenamine. During this time there is no significant increase in toxicity, and more elaborate precautions, such as the exclusion of air by covering the solution with a layer of mineral oil, are no longer considered to be necessary.

The injections of the standard being completed, the unknown is now weighed out and dissolved in the same way as the standard. The dose injected is 80 per cent. of the dose of the standard (0·4 ml. of 2 per cent. solution per 20-gm. mouse). The second group of mice is injected within 10 minutes of the preparation of the solution of unknown, both groups are set aside, and the mortality observed for 3 days. The result of the test is read as follows:

(*a*) Mortality from unknown *equal to, or less than* mortality from standard—*Sample passes.*

(*b*) Mortality from unknown *more than* mortality from standard—*Sample fails.*

(*c*) Mortality in *both* groups less than 10 per cent. or more than 90 per cent. of the mice—*Test must be repeated.*

If sufficient mice are not available for the test, it is permissible to use groups of 10 mice instead of 20 for a preliminary test. If the mortality from the unknown is *at least 2 mice less* than the mortality from the standard, the unknown sample passes the test, without any further experiment. If, on the other hand, the mortality from the unknown is greater than this, the test must be repeated on 2 further groups of 10 mice. The mortalities from the 2 tests are added together, and the results treated in the same way as in the test which uses 2 groups of 20 mice from the outset.

Method of the *British Pharmacopoeia*, 1948, for sulpharsphenamine. Sulpharsphenamine is not irritant to the tissues, and may therefore be injected subcutaneously without causing necrosis. The official test for absence of undue toxicity is carried out in exactly the same way as that described for neoarsphenamine, but the injections are given subcutaneously instead of intravenously. There is an additional test for absence of irritant activity.

The sample is dissolved in freshly distilled water to produce a 10 per cent. solution. Doses equivalent to 0·35 mg. of sulpharsphenamine per gm. of body-weight are injected subcutaneously into 5 rats, each weighing about 100 gm. (0·35 ml. of 10 per cent. solution per 100-gm. rat). No oedema or local necrosis should be produced in any of the rats for the sample to pass the test, and none of the rats should die. If more than 1 of 10 or 2 of 15 rats die after injection in the above manner, the sample fails to pass the test.

(*b*) *Test for therapeutic potency*

This test helps to ensure uniformity of product, but, as a measure of the activity of the drug in human therapeutics, it is of questionable value. It derives from Ehrlich's original work on salvarsan in which active organic arsenicals were selected upon the basis of their trypanocidal activity. Tests upon spirochaete and spirillum infections in animals, and syphilis in man followed. An investigation by workers at the National

Institute for Medical Research made in 1922, indicated that the trypanocidal activities of neoarsphenamine samples were proportional to their activity in removing spirochaetes from the chancres of human syphilis during the 18–20 hours following a single injection. All of the samples tested caused disappearance of the spirochaetes after 3 or 4 days, and the evidence presented does not prove that, in their ultimate effects, the less toxic and less trypanocidal samples were less effective than the more toxic and more trypanocidal samples. Much work has been done since this time in trying to produce samples of low toxicity and high trypanocidal potency. Further investigation is required, because it is possible that the standard set by the test for trypanocidal potency is not the best one, and that useful batches of drug are sometimes rejected because of their failure to comply with a test for activity upon the unrelated species of parasite.

The method specified in the *British Pharmacopoeia*, 1948, is based upon the work of Hawking (1943) which has been discussed already. The criterion of activity is the proportion of mice infected with *T. equiperdum*, in which the peripheral blood is cleared of trypanosomes by the injection of a single dose of the drug. Comparison with the standard preparation is made in the same test. The test is designed to exclude samples having a potency of less than 80 per cent. of that of the standard.

Method of the *British Pharmacopoeia*, 1948, for neoarsphenamine. A suitable suspension of *Trypanosoma equiperdum* is prepared from the blood of an infected rat, using saline containing 1 per cent. of sodium citrate as a diluent. Quantities of 0·5 or 0·25 ml. (well stirred), are injected into the peritoneal cavities of mice, the weights of which do not differ by more than 5 gm. The minimum number of animals required for the test is 20, and it is best to inoculate a few extra mice, so that a selection may be made.

After 48 hours (sometimes 24 hours is sufficient), the blood of each mouse is examined. The tip of the tail is snipped, and a drop of blood is transferred to a slide and pressed beneath a small coverslip. On examination under the ⅙th-inch objective of the microscope, the red cells should lie fairly close together, but in a single layer. The approximate degree of infection suitable for test is defined as 'one to twenty trypanosomes per 2 microscopic fields, of area approximately 0·12 sq. mm.; this

corresponds approximately to an infection of 1,000 to 20,000 trypanosomes per cu. mm. of blood'.

Two groups of 10 infected mice are selected so that the weights of the animals and their degree of infection are distributed as evenly as possible between the 2 groups. 0·1 per cent. solutions of the standard and unknown samples of neoarsphenamine are prepared as in the toxicity test, and the groups of mice injected intravenously within 10 minutes of preparation. The dose of the standard should be chosen from experience to be of sufficient size to cause disappearance of trypanosomes from the blood of about 50 per cent. of the mice. This quantity varies a little from one laboratory to another, but is usually about 0·24 ml. of 0·1 per cent. solution per 20-gm. mouse. The doses of the standard and unknown are given in the proportion of 0·8 to 1·0; thus if the group receiving the standard have 0·24 ml. per 20 gm., the group receiving the unknown have 0·3 ml. per 20 gm. The dose is also adjusted to the body-weight of each mouse.

On the first and third days following injection of the drug, a coverslip preparation of the blood of each mouse is examined. If no organisms are seen after the careful search of twenty fields under the $\frac{1}{6}$-inch objective, the mouse is recorded as 'cleared'. The maximum number of mice cleared by each sample of drug is determined. Mice which are free from infection on the first day, but dead on the third, are counted as cleared, but if more than 20 per cent. of the mice so die, the test must be repeated. This is because in such a case, some uncontrolled factor must be operating to cause the death of mice, once cleared of infection, in so short a time. If the dose-level of the standard has been chosen satisfactorily, an unknown sample of neoarsphenamine of potency equal to the standard would be expected to 'clear' all of the mice. The *Pharmacopoeia* allows a sample to pass if the number out of 10 mice cleared, exceeds by 2 or more the number out of 10 cleared by the smaller dose of standard. If this required proportion is not attained, the test is repeated on further groups of 10 mice, and the results of the two experiments combined. Then, if the total number of mice cleared by the unknown is equal to, or more than the number cleared by the standard, the sample passes the test. If there is still doubt, it is permissible to increase the number of mice in each group to a total of 40.

In the event of the clearance produced by *both* standard and unknown being less than 20 per cent. or more than 80 per cent. of the animals, the test must be repeated with more suitable dose-levels, because comparisons using the upper or lower extremes of the dose-response curve are unreliable.

Method for sulpharsphenamine. The details of the test for sulpharsphenamine are the same as for neoarsphenamine, except that:

(i) 0·2 per cent. solutions are used, because the trypanocidal potency of this drug is about one-half of that of neoarsphenamine.

(ii) Injections are given subcutaneously.

(iii) The mice are examined on the third and fourth days after injection, because sulpharsphenamine produces its maximum effect more slowly.

2. Antimonials

Sodium stibogluconate (B.P.C.). This drug is used for the treatment of leishmaniasis. A reference preparation is kept at the Pharmacological Laboratory, The School of Pharmacy (University of London).

The monograph in the *British Pharmaceutical Codex* requires that the drug shall be tested for absence of undue toxicity by the method described for neoarsphenamine in the *British Pharmacopoeia*, 1948 (see above). The solution used for the test is a 30 per cent. solution in water, which is injected intravenously into groups of mice. A dose of about 0·16 ml. is sufficient to cause death of about 50 per cent. of the animals. Sodium stibogluconate does not decompose in water, and there is no need to take special precautions during the preparation and injection of the solution.

Urea stibamine (*Indian Pharmacopoeial List*, **1946**). Urea stibamine is an amorphous mixture of derivatives of *p*-aminophenylstibonic acid and is widely used in India for the treatment of kala-azar. The monograph for the drug in the *Indian Pharmacopoeial List* specifies a toxicity test which is modelled upon the method of the *British Pharmacopoeia*, 1932, for neoarsphenamine.

A 2 per cent. solution of the sample is prepared in freshly boiled water, and is injected intravenously into mice weighing

13–15 gm. Each of 10 mice receives 200–250 mg. per kg. of body-weight. If not more than 2 mice die within 3 days of injection the sample passes the test. If more than 2 mice die, a second group of 10 mice is injected in the same way. If the total number of mice which die in the 2 tests is 8 or less, the sample passes the test. If the number which die is more than 8 and less than 15, a further group of 10 mice may be injected, and if the total deaths out of 30 mice is 15 or less, the sample passes. If the total is greater than 15 the sample fails to pass the test.

When a standard preparation becomes available, no doubt this test will be improved along similar lines to the neoarsphenamine test in the *British Pharmacopoeia*, 1948.

3. SURAMIN, B.P.

This substance, also known as 'Bayer 205', 'germanin', or 'antrypol', is used for the treatment of trypanosomiasis in man and animals. There is, as yet, no standard preparation available, and the biological tests specified in the *British Pharmacopoeia*, 1948, are similar to those for neoarsphenamine in the *British Pharmacopoeia*, 1932.

Test for absence of undue toxicity. A 2·5 per cent. solution of the substance is prepared in freshly distilled water, and each of a group of 10 mice is injected intravenously with 0·0125 ml. per gm. of body-weight (0·32 mg. of suramin per gm.).

If not more than 5 mice die within 3 days, the sample passes the test. If more than 5 mice die, a further group, this time of 20 mice, is injected. If the total deaths within 3 days of the injection are 15 or less, the sample passes the test. If the number of deaths is greater than 15, the sample is rejected.

Test for therapeutic potency. This is carried out in a manner similar to that described above for the standardization of neoarsphenamine. Mice are infected with *Trypanosoma equiperdum*, or other suitable species of trypanosome sensitive to suramin (which includes *T. rhodesiense*, *gambiense*, *brucei*, *evansi*, and *equinum*). After 48 hours the blood of each mouse is examined, and animals with approximately 1,000–20,000 organisms per μl. are selected. Ten mice are injected intravenously with 0·016 ml. of a 0·005 per cent. solution of the sample of suramin, prepared in freshly distilled water. The blood of each mouse is examined microscopically under a cover-

slip on the first and third days following. If no trypanosomes are found in the blood of 5 or more mice when 20 fields are examined under the ⅙-inch objective on the third day, the sample passes the test. If less than half of the mice are 'cleared', the test may be repeated, and the sample is allowed to pass if not less than 50 per cent. of the total number of animals are freed from trypanosomes in the peripheral blood.

REFERENCES

Browning and Calver, 1943, *J. Path. Bact.* **55**, 393.

Bülbring and Burn, 1938, *Quart. J. Pharm. Pharmacol.* **11**, 67.

Calver, 1945, Thesis for Ph.D., University of Glasgow (see abstract in *Trop. Dis. Bull.* 1945, **42**, 704).

Chen, Geiling, and MacHatton, 1945, *J. infect. Diseases*, **76**, 144.

Collier, Fulton, and Innes, 1942, *Ann. Trop. Med. Parasitol.* **36**, 137.

Durham, Gaddum, and Marchal, 1929, *Spec. Rep. Ser. Med. Res. Coun., Lond.*, No. 128.

Goodwin, 1944, *J. Pharmacol.* **81**, 224.

Gray, Trevan, Bainbridge, and Attwood, 1931, *Proc. Roy. Soc. B*, **108**, 54.

Hawking, 1941, *Quart. J. Pharm. Pharmacol.* **14**, 337.

—— 1943, ibid. **16**, 13.

Kolmer, 1926, *Principles and Practice of Chemotherapy*. W. B. Saunders & Co., Philadelphia.

The British Pharmacopoeia, 1932 and 1948. Constable, London.

The British Pharmaceutical Codex, 1949. The Pharmaceutical Press, London.

The Indian Pharmacopoeial List, 1946. Government of India Press, Calcutta.

Voegtlin and Smith, 1920, *J. Pharmacol.* **15**, 475.

Wien, 1946, *Brit. J. Pharmacol.* **1**, 65.

Yorke, Adams, and Murgatroyd, 1929, *Ann. Trop. Med. Parasitol.* **23**, 501.

Yorke and Murgatroyd, 1930, ibid. **24**, 449.

CHAPTER XXIII

AMOEBICIDAL SUBSTANCES

ALL who undertake chemotherapeutic tests with *Entamoeba histolytica* should first read the writings of Clifford Dobell (1919, 1928, 1929). This meticulous work forms the basis of most of our present knowledge of the morphology and life-history of the amoeba in man, monkey, and culture tube. For experiments upon the action of drugs the monkey is probably the ideal host, because it is found naturally infected with an amoeba which is indistinguishable in every way from *E. histolytica*. Dobell showed that this infection could be cured by emetine bismuth iodide given by mouth. More recent work upon the monkey infection has been carried out by Anderson, Johnstone, and Hansen (1947). Monkeys are difficult to house and handle, and a smaller laboratory animal is needed when large numbers of drugs are to be tested. When trophozoites of *E. histolytica* from culture or from human faeces are inoculated *per rectum* into kittens, severe infections, with diarrhoea and ulceration of the bowel are produced, which usually kill the animals in a few days. Dale and Dobell (1917) showed that such infections resisted treatment with emetine, but in spite of this, kittens have been used on a very large scale for the selection of new drugs for clinical trial (Wagner, 1935; Bieling, 1935; Kikuth, 1945). It is not surprising that the drug found to have most promise in the kitten test ('gavano') was without appreciable activity in man.

Rats are also natural hosts of *E. histolytica*, but early experimental infections gave very variable results. Jones (1946, 1947) showed that when amoebae were injected directly into the caeca of weanling rats, heavy infections with ulceration of the gut wall were produced in 60–100 per cent. of the animals. Also, this infection was prevented from developing by giving doses of emetine or chiniofon. The young rat is probably the most useful small animal at present available for the study of amoebiasis.

Maintenance of strain. Amoebae are maintained in the laboratory in media containing serum and rice starch. A large

number of different modifications have been described, and the one to choose is that which is found to give the best results in the hands of the worker concerned. Jones (1946) uses a modification of Pavlova's medium, with the composition:

Sterile horse serum	. 0·5 ml.
1 per cent. marmite solution	. 1·0 ml.
Buffered saline (pH 7·2)	. 8·5 ml.
Rice starch	. 30 mg.

The buffered saline is made by mixing:

Disodium hydrogen phosphate (9·476 gm. per litre)	. 375 ml.
Potassium dihydrogen phosphate (9·078 gm. per litre)	. 125 ml.
Sodium chloride (9·0 gm. per litre)	. 2,250 ml.

Alternatively, the media described by Dobell and Laidlaw (1926) may be used, a good one being the solid horse serum slope overlaid with a filtered solution of egg-white in Ringer, and with the addition of some solid rice starch.

When isolating new strains of amoebae from human faeces, it is advisable to inoculate tubes of a number of different types of medium, and to choose the formula which best suits the particular strain isolated. The addition of acriflavine, 1 part in 10,000 of the medium, during the first few passages helps to prevent the protozoa from being swamped by a luxuriant growth of micro-organisms, particularly starch-splitters and *Blastocystis*. Acriflavine at this concentration has no action on the amoebae. The tubes are incubated at 37° C. and must be passaged twice a week. About 0·1 ml. of the starchy deposit, in which the amoebae grow, is carefully removed with a Pasteur pipette, and transferred to the surface of the starch at the bottom of a clean tube of medium. The least disturbance of the fluid part of the medium, the better. Tubes of culture must not be allowed to become chilled on the bench, but must be replaced as soon as possible in the incubator.

Preparation of inoculum. For the inoculation of rats, large numbers of amoebae are required. They may be grown in bottles (300 ml.) of the fluid medium described above, or in conical flasks containing Dobell and Laidlaw's medium (Frye and Meleney, 1935). Jones prepared the inoculum by taking the combined deposits from twelve 300-ml. bottles of medium, centrifuging, adjusting the volume to 8 ml. and adding an equal

volume of 10 per cent. solution of gastric mucin in order to thicken the inoculum. Such a suspension contains about 1 million amoebae per μl., and is sufficient for the inoculation of 60 rats.

Successful results may be obtained by using much smaller amounts of inoculum, and as the amoebae are fragile, it is doubtful if the process of centrifuging is necessary or desirable. The careful removal of the deposit with a Pasteur pipette into a warmed bottle is sufficient. There is no need to add mucin.

Rats. The rats used are weanlings of 17–21 days old, weighing 20–33 gm. They must be available from a breeding colony close by, because at this age they do not travel well, and easily fall victims to chilling or chance infections. It is important to use rats immediately they are weaned because then the caecum is a small pink organ containing only mucus and a very few protozoa and bacteria. In such an environment, an inoculum of amoebae has the best chance of producing a successful infection. If the rats are allowed to live for a day or two upon a solid diet before inoculation, the caecum rapidly becomes distended with a mass of food residues, bacteria, and protozoa. The introduction of *E. histolytica* into this mass is less likely to give a severe infection and, although amoebae may remain present in the caecal contents for many weeks, ulceration of the gut-wall rarely occurs.

Another important point is that the rats must be bred under scrupulously clean conditions, and must be bedded on wood-wool which has been sterilized by heating to 150° C. for 1 hour. This is because rats sometimes harbour an amoeba of their own, *E. muris*, which is readily acquired from bedding or food which has been fouled by wild rats or mice. Hamsters also frequently carry *E. muris* and should not be kept in the same room with rats used for experiments with *E. histolytica*. *E. muris* is non-pathogenic and is morphologically similar to *E. coli* of man. It is unaffected by emetine and other amoebicidal substances, and, unless mature cysts are present, is not easy to differentiate from *E. histolytica* in a fresh preparation. It is worth while taking especial care over the breeding and housing of the animals in order to avoid errors which may arise from the presence of this organism.

Method of inoculation. Each rat is anaesthetized with ether,

and the abdomen is shaved and swabbed with a 1 per cent. solution of cetrimide. Aseptic precautions should be taken, and a small glass-topped operating board such as that described in Chapter XXIV will be found convenient. A small cut, 0·5–1 cm. long is made with scissors, slightly to the left of the midline, and the caecum delivered. In the weanling rat the caecum is often found in the left iliac fossa, but the position is very variable. About 0·2 ml. of the suspension of amoebae is taken in a warm tuberculin syringe fitted with a 23 S.W.G. needle, and is injected through the wall of the caecum, so that the white material is seen to distend the organ and to make good contact with its walls. It is a good thing to enter the caecal wall at an angle, to avoid leakage of the inoculum into the peritoneum. The caecum is replaced and the wound secured with a stitch and a Michel clip. A little sulphanilamide powder may be dusted into the peritoneal cavity if desired, but is not necessary if reasonable care is taken. The rat's peritoneum is remarkably resistant to infection. The animals are put in a warm place to recover, and are then distributed into groups for the test.

Rats may be infected with amoebae directly from human faeces. If the specimen contains trophozoites, the injection is made into the caecum by the method described above. If cysts only are present, the suspension of faeces must be given by stomach tube; the cysts hatch in the small intestine, yielding trophozoites which establish themselves in the caecum.

Treatment with drugs. Treatment is begun immediately, or on the day after inoculation. Doses may be given by injection, by stomach tube, or by incorporation in the diet. If the drug-diet method is used, watch must be kept to see how much is consumed. A weanling rat will eat about 4 gm. of diet a day, but less is eaten if the drug present gives the food an unpleasant taste.

One group of animals is kept untreated, as a control. Groups of at least ten rats should be used for each substance tested.

Assessment of result. Ulceration of the caecum begins within 24 hours of inoculation and increases during the following 5 days. The degree of infection then begins to decrease in some animals, and with some strains of *E. histolytica,* the condition of the caecum often becomes almost normal in the course of 3 or 4 weeks. The best time to assess the effect of treatment is

therefore at about the sixth day after inoculation. The rats are killed, and the caeca removed, opened, and any faecal matter carefully wiped away. The amount of ulceration present is noted, and the mucosa is brushed very carefully with a wire or a glass rod and the scraping transferred to a drop of saline on a slide. This is covered and examined under the microscope, and the concentration of amoebae present recorded. A score is awarded according to the condition of each caecum:

Caecum apparently normal, no amoebae seen in smear . .	0
Caecum apparently normal, small number of amoebae . .	1
Caecum apparently normal, moderate number of amoebae .	2
Caecum slightly inflamed, numerous amoebae	3
Marked inflammation, mucus, and numerous amoebae . .	4
Ulceration, mucus, and numerous amoebae	5

The average degree of infection for the group of rats is calculated from the individual scores, and the figure compared with that found for the untreated controls by means of a significance test (Chapter III, Section III).

When a drug causes irritation of the gut, it is sometimes difficult to assign a score upon the basis of the above scale. For example, with large doses of emetine, although no amoebae are present, the wall of the caecum is frequently oedematous and the contents mucoid. Goodwin, Hoare, and Sharp (1948) used the percentage of rats freed from infection as the criterion of effect. In Fig. 75 a set of results is expressed by both methods of assessment.

In vitro Method

In spite of determined attempts, particularly in the U.S.A., it has not yet been possible to grow *Entamoeba histolytica* free from bacteria *in vitro*. Cultures prepared from human faeces contain a mixed flora which make the results of chemotherapeutic tests impossible to interpret with certainty. A drug added to the medium may kill amoebae by direct action, or by killing one or more species of bacterium acting as sources of essential nutrients. Chinn, Jacobs, Reardon, and Rees (1942) have isolated cysts by micro-methods and have grown them with single strains of bacteria. The most successful cultures were obtained with a *Clostridium* designated 'organism *t*'. There is now a great deal of information available upon the requirements and properties of this pure mixed culture. In

this country, Dobell has prepared a pure mixed culture of *E. histolytica* with a smooth strain of *Bact. coli* by a method which has not yet been published. Cultures of this kind are satisfactory for *in vitro* tests of amoebicidal substances, because any action of the drug upon the bacteria present may be studied, and taken into account.

Dobell (1947) described a method of conducting the test upon

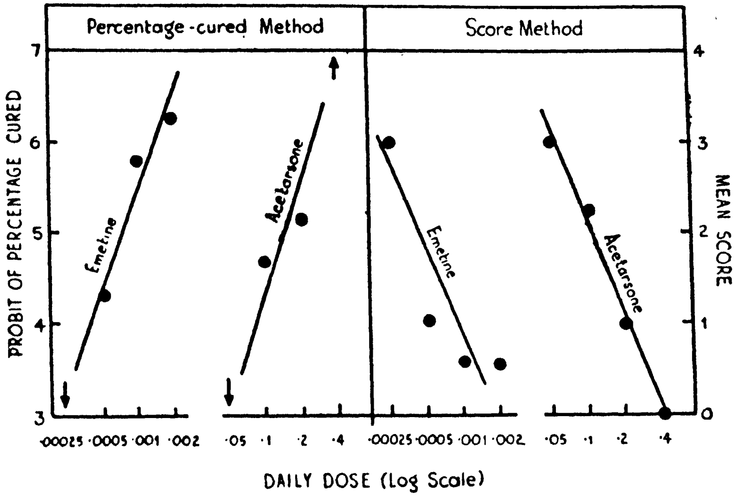


FIG. 75. Results of a test upon emetine hydrochloride and acetarsone given by mouth as drug-diet, plotted as percentages of animals protected and as scores awarded according to the condition of the caecum.

his pure mixed culture which was used for the investigation of the activity of substances allied to emetine by Goodson *et al.* (1948).

A fluid medium consisting of one part of inactivated horse serum with eight parts of Ringer to which 0·2 per cent. of disodium hydrogen phosphate has been added, is distributed into test-tubes in 5-ml. quantities. Serial dilutions of drugs are prepared in the usual way: some tubes are kept as controls. A large loopful (about 20 mg.) of dry, sterile rice starch is added to each tube, followed by one drop of a 1 per cent. solution of methylene blue, to act as an indicator. Methylene blue has no effect upon *Bact. coli* or amoebae, but if it interacts with the drug to be tested it should be omitted from all except the control tubes.

About 5 drops of a rich 24-hour culture of the *Bact. coli* in Douglas-Hartley broth is added to each tube and all are incubated over-night at 37° C. If satisfactory growth of bacteria occurs the dye will be bleached and the conditions at the bottom of the tube will be favourable for the growth of amoebae. If the drug has inhibited the growth of the bacteria, it will be apparent at this stage. The deposit from a rich culture of the pure mixed strain of amoebae and bacteria is then taken, and used to inoculate all of the tubes. A Pasteur pipette is charged with the suspension, which is deposited with as little disturbance as possible at the bottom of each tube. After 48 hours in the incubator, a vigorous growth of amoebae will be observed in the control tubes, and the concentration of an active drug which will just inhibit growth can be determined. Emetine hydrochloride tested by this method is active at a concentration of 1 : 5,000,000.

Every batch of serum used for the preparation of medium should first be proved in preliminary tests, because all are not equally good. Poor samples may be improved by the addition of 0·1 per cent. glucose to the Ringer's fluid used to dilute the serum.

REFERENCES

Anderson, Johnstone, and Hansen, 1947, *Am. J. Trop. Med.* **27**, 153.
Bieling, 1935, *Beih. Arch. Schiffs-u. Tropenhyg.* **39**, pt. ii, 7.
Chinn, Jacobs, Reardon, and Rees, 1942, *Am. J. Trop. Med.* **22**, 137.
Dale and Dobell, 1917, *J. Pharmacol.* **10**, 339.
Dobell, 1919, *The Amoebae Living in Man.* John Bale, Sons and Danielsson, London.
—— 1928, *Parasitology,* **20**, 357.
—— 1947, *Ann. Soc. Belge Med. Trop., Liber Jubilaris J. Rodhain.*
Dobell and Bishop, 1929, *Parasitology,* **21**, 446.
Dobell and Laidlaw, 1926, ibid. **18**, 283.
Frye and Meleney, 1935, *Science,* **81**, 99.
Goodson, *et al.*, 1948, *Brit. J. Pharmacol.* **3**, 49.
Goodwin, Hoare, and Sharp, 1948, ibid. **3**, 44.
Jones, 1946, *Ann. Trop. Med. Parasitol.* **40**, 130.
Kikuth, 1945, *B.I.O.S. Rep.* 116, item 24, p. 30. H.M. Stationery Office, London.
Wagner, 1935, *Beih. Arch. Schiffs-u. Tropenhyg.* **39**, pt. i, 9.

CHAPTER XXIV

LEISHMANICIDAL SUBSTANCES

THE introduction of tartar emetic in 1913 by Vianna for the treatment of leishmaniasis, and the development of the less toxic quinquevalent antimonials is a well-known story in the history of chemotherapy. However, it was not until 1924 that a satisfactory laboratory infection was available, and the first drug (solustibosan) to be introduced because of its activity against kala-azar in the hamster, did not appear until 1937. The most widely used antimony compounds, urea stibamine, neostibosan, and neostam, were all developed as a result of experiments upon mouse trypanosomiasis and clinical trials. How many useful drugs were rejected during this period it is impossible to compute, because low activity against *Trypanosoma equiperdum* in the mouse is not always paralleled by low activity against the leishmanias. Solustibosan is a good example of a drug which has very little trypanocidal action and is yet a potent leishmanicide.

The infection of rats and mice with leishmanias is uncertain and unsatisfactory. Dogs, which play the part of a reservoir host in Mediterranean countries, are inconvenient for laboratory tests. Smyly and Young (1924) discovered that the Chinese hamster *Cricetus griseus* developed a severe and lethal infection when inoculated with infected human material, and since that time a large variety of rodent species have been shown to be susceptible (e.g. *Meriones*, *Sigmodon*, *Citellus*). Gerbils (*Rhombomys*) have been demonstrated to be a natural reservoir host of oriental sore in Soviet Central Asia. The most commonly used laboratory species is the golden hamster *Cricetus auratus*, which breeds well in captivity and is of a convenient size. This attractive little beast has long and sharp incisors, and powerful hind legs, but it does not bite if handled gently, and is becoming increasingly popular as a domestic pet.

Röhl (1929) described a laboratory test in which hamsters were infected and then treated with antimony compounds. He showed that the infection was delayed, or failed to develop when active drugs were injected. The method was based upon

his well-known test for antimalarial drugs. The test was elaborated by Wang (1938), Kikuth and Schmidt (1938, 1941), Adler and Tschernomoretz (1939, 1941, 1942), Goodwin (1944, 1945), and Fulton (1944), and is now firmly established as a screening test and method of assay. It has been responsible for the introduction of the aromatic diamidines for the treatment of leishmaniasis.

The Life-cycle of the Leishmania Parasite

In the mammalian host, the parasite lives and multiplies within the cells of the reticulo-endothelial system, which are stimulated to increase in number and to form large areas of 'clasmatocyte tissue' in the spleen, liver, lymph glands, skin, and other organs. Very few parasites appear in the blood, where they are found within the leucocytes. The spleen and liver enlarge, amyloid degeneration may develop in the kidneys, and the animal wastes and usually dies from a terminal pneumonia or other secondary infection. The parasite is transmitted in nature by the sandfly, *Phlebotomus*, in which the small, rounded leishmanial form exflagellates and becomes a leptomonad. After a suitable time, the flagellates multiply and become so numerous as to block the fore-gut of the insect. Some of them become dislodged and are injected into a second mammalian host in the sandfly's efforts to suck blood into its occluded alimentary tract.

Maintenance of strain. The flagellate form may be cultured in the laboratory in a medium containing blood such as 'N.N.N.' or Adler's Locke-blood-agar, kept at a temperature of 22–30° C. The organisms die if kept at 37° C. Sub-inoculations should be made at intervals of 1–3 weeks. Hamsters may be infected by the repeated intraperitoneal inoculation of flagellates, but the infection develops more rapidly if the injection is made directly into the spleen under anaesthesia with ether or a barbiturate. The spleen is exposed through a small abdominal incision, about 0·1 ml. of a heavy suspension of flagellates injected, and the wound is closed. The infection develops in from 1 to 6 months, depending upon the strain of leishmania. For routine passage and the inoculation of animals for test, it is more satisfactory to inject a suspension of the mammalian forms, made by grinding up the spleen and liver of a heavily infected hamster. A

single intraperitoneal inoculation with a virulent strain produces heavy infections in 4–6 weeks.

The spleen increases in size, mainly as a result of engorgement with blood, but later, when the infection is very severe, frequently contracts to almost normal proportions and becomes pale in colour. The size of the spleen is therefore no index of the degree of infection.

Estimation of Degree of Infection

It is possible to obtain an approximate idea of the degree of infection in a hamster by liver puncture with a large-bore needle. The method needs some practice and is not entirely satisfactory. The amount of reticulo-endothelial tissue in the liver is not great, and granular material from the liver parenchyma may obscure the parasites in stained smears prepared from the aspirated tissue. It is more satisfactory and very little more trouble to remove a small fragment of spleen under anaesthesia, and to prepare dab-preparations in which all of the cell-types present and the parasites are easily identified and counted when stained with Giemsa's stain. If aseptic precautions are taken, the mortality resulting from spleen biopsy is negligible.

The degree of infection may be estimated approximately by eye as '+', '++', '+++', or '++++', or by a method in which the parasites are counted. Fulton (1944) prepared smears as nearly as possible of the same density, and counted the number of parasites seen in 20 microscope fields under the oil-immersion objective. Adler and Tschernomoretz (1939) counted the number of parasites present per 100 spleen cell nuclei of all types, and this method has been shown by Goodwin (1944) to give a reliable estimate of the degree of infection present throughout the spleen. The number of parasites and nuclei which must be counted to give a reliable figure depends, of course, upon the intensity of the infection, but it is usually adequate to count 250–500 spleen cell nuclei and the accompanying parasites.

Treatment with Drugs

For a 'screening' test upon an unknown drug, doses approaching the toxic level are given once or twice daily for a period of 1 or 2 weeks, beginning on the day after the initial spleen

biopsy. The biopsy is repeated, or the animals are killed and examined, after a further period of a week to give time for the resolution of parasites killed by the treatment. Some workers give doses of drug twice a week for 4–5 weeks or longer. A group of infected hamsters should always be kept as controls; they may afterwards be used as a source of parasites for further tests. A simple test such as this will pick out any substance likely to be of interest as a leishmanicide, because if the drug has had any effect the degree of infection will be reduced. The infection in control animals always increases (often very rapidly), after the initial biopsy.

For the assay of antimony compounds, it is usually sufficient to give a single dose of the drug and to re-examine the spleen 1 week later. The aromatic diamidines are more slow in their action, and it is best to give a course of injections extending over 3 or 4 weeks before assessing the effect of the treatment. Injections are best given by the subcutaneous or intramuscular route. Intravenous injection in the hamster, which has only a short stump of a tail, must be made under anaesthesia into the jugular or femoral vein.

Assessment of the Effect of a Drug

For screening tests it is sufficient to note if treatment with a drug reduces the number of parasites present in the spleen. For the assay of drugs known to be active, the method of assessment of activity used most widely has been to give doses by trial and error until the relative activities of the compounds can be estimated by inspection of the results. Wang (1938) in comparing the effects of solustibosan and neostibosan, tried to find the dose of drug which would cure the animals, and went so far as to test for cure by sub-inoculation of the tissues of treated animals into fresh hamsters. In the ordinary way this method (which takes, in all, 6–12 months to complete) is unnecessary and it is sufficient to estimate the effect of smaller doses of drug in reducing the intensity of the infection. It has been shown (Goodwin, 1945) that the degree of infection in the spleen determined by Adler's method before and after treatment can be used to construct a linear dose-response curve. The number of parasites per 100 spleen cell nuclei is determined in smears made from biopsied material before and after treatment,

and the initial count is divided by the final count to give a ratio which is large if the effect of the drug has been considerable, and unity if there has been no change in count. The mean of the logarithms of the ratios determined for a group of animals is plotted against the logarithm of the dose at each dose-level, and a straight line results. By plotting the dose-response curves of two drugs in this way, their relative activities may be compared, graphically, or by calculation.

The counting of parasites in smears is very tiresome, but for the accurate determination of the potency of drugs of the same chemical type, and for the calculation of the limits of error of the results, it is the only method so far available.

Experimental Details

(*a*) **Inoculation of animals.** A heavily infected hamster is killed with ether, the spleen is removed with aseptic precautions, and separated from mesentery and fat. It is then rapidly ground in a mortar to form a smooth paste and diluted with normal saline to a suitable volume—usually about 50 ml. A heavily infected spleen containing 100–500 leishmanias per 100 spleen cell nuclei is sufficient to inoculate about 100 hamsters. If desired the liver may also be added to the inoculum. The suspension is strained through gauze and injected into hamsters, 0·5 ml. being given by intraperitoneal injection. The suspension must be kept stirred, and the injections should be completed as soon as possible after preparation of the inoculum.

If the strain of leishmania is a virulent one, the animals will have developed a sufficiently heavy infection in 4–6 weeks after inoculation. During this time the animals should preferably be kept in individual cages, but they come to little harm when kept in groups of 5–10 if the sexes are segregated and the cages are sufficiently large.

(*b*) **Method of spleen biopsy.** After 4–6 weeks, a few hamsters are examined to see if the infection in the spleen has become sufficiently heavy. If so, the rest of the group is examined without delay.

The animals are anaesthetized one by one by the intraperitoneal injection of sodium hexobarbitone 100 mg. per kg., made up freshly as a 2 per cent. solution in distilled water. Alternatively ether may be used, but the animals require more

attention during the operation. Pentobarbitone is not suitable for hamsters. The anaesthetized animal is stretched by strings attached to the left fore and hind limbs upon a small operating board. A suitable board may be made from a small frame 12×8 in. in size, with a detachable glass top and with cleats screwed to the upper and lower edges. Such a board is convenient and is easily kept clean. The left side of the abdomen is swabbed with a 1 per cent. solution of cetrimide (C.T.A.B.), shaved, and covered with a small sterile gauze towel. An incision about 1 cm. long is made slightly to the left of the midline and midway between the pelvis and the costal margin. The lower pole of the spleen is easily found and a small fragment about 1 mm. long is snipped from the tip with dry sterile scissors. Bleeding is slight and is easily controlled by dabbing with a small cotton pledget; there is no need to apply ligatures. The spleen is replaced and the abdominal muscles closed with a thread suture. The skin is closed with a Michel clip (size, 1 cm.), which the hamster is unable to bite out. Other types of suture are usually quickly torn away as soon as the animal recovers from the anaesthetic.

The cut surface of the spleen fragment is wiped free of blood and then dabbed upon the surface of a clean dry slide. If the scissors and forceps used for the handling of the fragment are wet, distortion of the thin layer of cells and parasites occurs and the preparation is spoiled. The slides are fixed for 2 minutes in methyl alcohol, drained dry, and stained for 1 hour with Giemsa's stain, freshly diluted 1 drop in 2 ml. of water, the pH of which has been adjusted to 7·4 by the addition of a little lithium carbonate. Alternatively a phosphate buffer at pH 7·4 may be used for dilution of the stain.

The smears are examined under the oil-immersion objective and the animals roughly classified as light, medium, or heavy in infection. After spleen biopsy, the animals must be kept in individual cages or fights and deaths will occur. They are grouped so that each group contains approximately equivalent degrees of infection, and drug treatment is begun on the following day. If left for a day or two before giving treatment, the estimate of the initial degree of infection will be much too low, because the parasites often multiply very rapidly after the operation. There is not time to make full counts of the

proportion of parasites to spleen cell nuclei before beginning treatment. Detailed counts may be made during the days which follow, when the animals are receiving drug treatment.

At the end of the test a second biopsy may be made, or, if it is known that the drug will have exerted its action and no further record from the test is required, the animals may be killed and smears prepared from the spleens for the second parasite count.

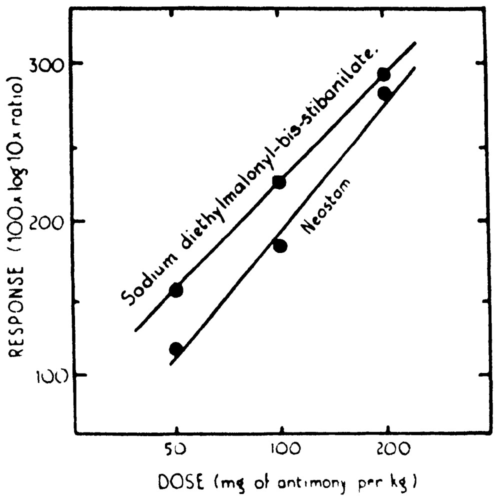


FIG. 76. Dose-response curves from the data in Table LXXXII. Single doses of the drugs were given. The logarithms of these doses are plotted against the logarithms of the ratios of parasite counts before and after treatment.

In Table LXXXII and Fig. 76 the results of a test are recorded. The relative potency of the drugs is calculated by the method described in Chapter III, Section VI (particularly § 59), or by the method of the British Standards Institution (1940).

In vitro METHODS

The flagellate form of *Leishmania donovani*, as stated above, is easily grown in the test-tube, and has therefore been used for the investigation of drugs for use in the treatment of leishmaniasis. Like other *in vitro* methods, it has limited uses, as is shown by the fact that quinquevalent antimony compounds have very little effect upon the organism in culture. Nevertheless,

TABLE LXXXII. *Example of calculation of the mean response to a dose of neostam*

To avoid fractions and negative indices with low doses, the response is taken as 100 times the logarithm of 10 times the ratio of the parasite counts before and after treatment.

Drug	*Parasites per 100 spleen cell nuclei*		*Log of ratio* $\frac{\textit{count before}}{\textit{count after}}$	*100 × log 10 × ratio*	*Mean response*
	Before treatment	*After treatment*			
Neostam	50	12·5	0·602	160	
(100 mg.	120	14·8	0·909	191	
of Sb.	56·5	20·1	0·449	145	
per kg.)	46·5	9·7	0·680	168	177·7
	100	18	0·745	175	
	59	7	0·926	193	
	160	12·1	1·121	212	

TABLE LXXXIII. *Mean responses to doses of neostam and sodium diethylmalonyl*-bis-*stibanilate*

Drug	*Dose mg. Sb/kg.*	*No. of hamsters*	*Mean response*
Neostam	50	8	116·5
	100	7	177·7
	200	8	288·4
Sodium diethylmalonyl-*bis*-stibanilate	50	8	154·9
	100	9	223·6
	200	6	291·8

Calculation by the method of the British Standards Institution (1940) gives:

1 mg. sodium diethylmalonyl-*bis*-stibanilate is equivalent to 1·30 mg. of neostam.

Limits ($P = 0{\cdot}95$): 76–134 per cent.

a useful study of the diamidines has been made by Collier and Lourie (1946) who showed that when the cultures were grown at as high a temperature as possible (34° C.), the relative activities of this series of drugs were in approximately the same order as when assayed against the infection in the hamster.

Method of Collier and Lourie

Preparation of medium. Rabbit blood is drawn aseptically from a vein, or by heart puncture, and is defibrinated. One

volume of blood is laked with six volumes of sterile distilled water, and one volume of a sterile 6·3 per cent. solution of sodium chloride added to readjust the tonicity. The solution is centrifuged to remove the red cell 'envelopes', and mixed with half its volume of fresh, unheated rabbit serum. This medium, dispensed into small test-tubes, inoculated with leishmania and kept at 24° C., produces a heavy growth of flagellates in 6–8 days.

Details of the test. The test is carried out in a manner similar to the *in vitro* trypanocidal test. The medium is used to prepare serial twofold dilutions of the drugs to be tested in Kahn tubes, each tube containing 0·45 ml. of the medium. The tubes are covered with little glass caps. A suspension of flagellates is prepared by first lightly centrifuging an 8-day culture to remove large masses of flagellates. The supernatant fluid containing free-swimming organisms is stirred, the flagellates counted in the haemocytometer, and the concentration adjusted by dilution or by further centrifuging, to give a suspension containing 1,400–1,700 organisms per μl. Quantities of 0·025 ml. of the suspension are added to each tube with a teat pipette.

After addition of the leishmanias the tubes of medium are incubated at 34° C. for 5 days. A few tubes containing pure medium without drug are used as controls. A drop from each tube is examined under the microscope, and the lowest concentration of drug in which parasites are not found is taken as the leishmanicidal titre of the substance.

REFERENCES

ADLER and TSCHERNOMORETZ, 1939, *Ann. Trop. Med. Parasitol.* **33**, 313.

—— —— 1941, ibid. **35**, 9.

—— —— 1942, ibid. **36**, 11.

COLLIER and LOURIE, 1946, ibid. **40**, 88.

FULTON, 1944, ibid. **38**, 147.

GOODWIN, 1944, *Trans. R. Soc. Trop. Med. Hyg.* **38**, 151.

—— 1945, ibid. **39**, 133.

KIKUTH and SCHMIDT, 1938, *Arch. Schiffs- u. Tropenhyg.* **42**, 189.

—— —— 1941, *Z. Immunforsch.* **100**, 157.

RÖHL, 1929, *Indian Med. Gaz.* **64**, 563.

SMYLY and YOUNG, 1924, *Proc. Soc. Exp. Biol. Med.* **21**, 354.

WANG, 1938, ibid. **39**, 418.

British Standard Method for the Biological Assay of Vitamin D_3 by the Chick Method, 1940. British Standards Institution, London.

CHAPTER XXV

ANTIMALARIAL SUBSTANCES

THE discovery of the antimalarials mepacrine and pamaquin was made in the laboratories of the I.G. Farbenindustrie by testing a series of synthetic substances against infections of *Plasmodium relictum* in canaries. Early in the Second World War it was discovered that several sulphonamides, such as sulphadiazine and sulphamerazine, had considerable activity against human malaria, but were practically inactive when tested against the canary infection. Also, it was realized that the canary test, as devised by Roehl, would detect only activity against the cycle of schizogony in the erythrocytes, and that it was of great importance to find drugs which had a true prophylactic action.

Since that time many thousands of drugs have been tested for activity in experimental malaria, and the story of this intensive programme of research has been told elsewhere (Burn, 1948; Wiselogle, 1946). The experimental infections most widely used were those of *P. gallinaceum* in chicks and *P. lophurae* in ducklings. Canaries infected with *P. relictum* or *P. cathemerium* and monkeys infected with *P. knowlesi* or *P. cynomolgi* were also used. Drugs were tested against as many different species of parasite as possible, to minimize the chance of missing an active substance.

P. gallinaceum in the chick is the most convenient infection to use. It is sensitive to all drugs so far discovered which have activity in human malaria. Chicks are cheap, easy to obtain and to keep in the laboratory, and their small size enables tests to be made upon drugs which are in limited supply. The technique of the test is simple, and a clear answer is usually obtained in 1 or 2 weeks. The test can be designed to detect activity upon the merozoites, or to detect prophylactic activity against infections produced by the inoculation of sporozoites.

Principle of the Test using *P. gallinaceum* in Chicks

When infected blood is inoculated intravenously into a 6-day-old chick merozoites are to be found in the circulating

blood immediately after injection. The number of parasited erythrocytes increases rapidly to a maximum in 4 or 5 days, when 50–90 per cent. or more of the circulating red cells contain schizonts. Gametocytes also make their appearance. Some chicks die (probably from anaemia) and the intensity of infection in the blood of those that survive then begins to wane. Exoerythrocytic forms develop in the cells of the reticulo-endothelial system, and the rest of the chicks die 14–21 days after inoculation from blockage of the brain capillaries by huge multinucleated parasites (James and Tate, 1938). If the chicks used are more than 10 days old a number of them survive, and although parasites may disappear from the blood the organisms may be recovered from the brain, liver, and spleen for long periods.

When the infection is induced by the inoculation of sporozoites, either by a mosquito or by means of a syringe, the sporozoites first enter reticulo-endothelial cells and undergo development as the 'cryptozoite' or 'primary tissue phase'. For a period of 24–36 hours no parasites can be detected in the blood-stream, either by microscopical examination or by subinoculation. Trophozoites then begin to appear in the erythrocytes, and the cycle of schizogony proceeds. Large secondary tissue forms also develop from cryptozoites or trophozoites, and most of the birds die from blockage of the brain capillaries 9–12 days after inoculation of the sporozoites. If drugs are given before the appearance of merozoites, action upon the primary tissue phase may be detected, because then no merozoites appear and no secondary tissue forms are found when the chicks are killed. If the drug is active against the merozoites, but not against the primary tissue forms, the blood infection may be suppressed, but exoerythrocytic forms survive and multiply. No drug has yet been found which has activity against sporozoites.

Descriptions and illustrations of the cycle of *P. gallinaceum* in the chick will be found in the paper by Huff and Coulston (1944) and in the book by Hewitt (1940). Full details of the methods of test used by the different laboratories in the United States for the screening of antimalarial substances are given in the official report (Wiselogle, 1946). The methods described below are those of Curd, Davey, and Rose (1945).

Blood-induced Infections of *P. gallinaceum* in Chicks

Chicks. Day-old chicks from a reliable source and of the same 'cross' are kept for a week in the laboratory in brooders. The temperature under the hover should be 90° F. and the temperature of the room should be not less than 70° F. Male birds are cheaper than female, and if one of the standard sex-linked colour crosses such as Rhode Island × Light Sussex is used, the uniformity of strain is also ensured. The chicks are fed upon a dry diet such as 'National Chick Meal', which is given *ad lib.* Clean, preferably running water must also be supplied. At the age of 6 days the chicks weigh from 40–55 gm., and are suitable for inoculation with plasmodia. All weakly and underweight birds are rejected.

Preparation of inoculum. A blood smear is prepared from a chick infected not more than 5 days previously, and is stained with Leishman or Giemsa stain. A convenient method of obtaining blood is to prick the vein which runs on the medial side of the ankle with a sharp needle: a haematoma rarely forms in this situation as the surrounding tissue is dense. The smear is examined under the oil-immersion objective, and the percentage of erythrocytes which contain malarial parasites is determined. For the preparation of a satisfactory inoculum, infection of more than 50 per cent. of the red cells is required. No notice need be taken of the size of the schizonts in the individual red cells. If the parasites are almost all ripe schizonts, the infection produced in the inoculated birds will mount more rapidly than if the inoculum contains mostly young schizonts and ring-forms, but in either case the peak of the infection occurs within 4 or 5 days.

The chick is anaesthetized with ether, and the feathers damped with antiseptic. The heart is exposed by cutting along both sides of the cartilaginous sternum with scissors, and the blood is removed from the right ventricle with a Pasteur pipette containing a little anticoagulant solution. Chick blood clots very easily, and must therefore be transferred without delay into a tube containing about 2 ml. of 1 per cent. sodium citrate, or 0·1 per cent. heparin. Davey uses a mixture of citrate and heparin, which is more efficient than either alone. Clotting is also delayed by drawing the blood into glassware

lined with a film of paraffin wax or of a plastic such as 'Perspex' or polythene.

The number of red cells per μl. of the suspension is determined by counting in the haemocytometer chamber, and from this figure and that obtained from the stained smear, the number of parasited erythrocytes per μl. of the inoculum is calculated. The inoculum is diluted with anticoagulant so that every 0·2 ml. contains about 50 million parasited cells.

Method of inoculation. The inoculum is well stirred, and is drawn into a tuberculin syringe fitted with a 23 S.W.G. needle. The chick to be inoculated is held supine in the palm of the right hand of an assistant, with the forefinger behind the bird's neck. The operator takes the head between the thumb and index-finger of the left hand, and stretches the neck gently but firmly over the assistant's forefinger. The assistant swabs the neck with a pad well charged with cetrimide, and parts the feathers to display the right jugular vein: he places his left forefinger over the vein at the base of the neck so that it becomes distended with blood. The vein is rather mobile, but with very little practice is easy to enter with a sharp needle. 0·2 ml. of the inoculum is injected, and the point of entry pinched for a moment to avoid the production of a haematoma—which very easily forms in this situation.

The inoculated chicks are weighed and numbered with leg or wing-bands, and put into cages in groups of six. They are kept in a room maintained at a temperature of 80° F. In most parts of the world mosquitoes are to be found which are capable of transmitting *P. gallinaceum*. For the protection of poultry flocks it is therefore desirable that all animal rooms in which infected birds are kept should be screened with fine mosquito netting, and should have a system of double doors so that all insects can be prevented from entering the room.

Treatment with drugs. Drugs are usually given by mouth. Marshall, Litchfield, and White (1942), who tested drugs upon *P. lophurae* infections in ducklings, gave the drug in the diet. In order that a steady level of the substance should be maintained in the blood these workers fitted the cages with artificial lighting operated by a time clock, so that periods of 3 hours of light alternated with 3 hours of darkness. Birds eat in the light and sleep in the dark. Such an installation is unnecessarily

complicated for most purposes, because the birds are very easy to dose by mouth, using a short piece of gum-elastic or rubber catheter, or a large blunted needle attached to a syringe. The chick is held in the palm of the left hand and the neck extended by holding the head between the thumb and index-finger. The syringe is held in the right hand and the catheter or needle introduced at the angle of the mouth. It passes easily and smoothly down the gullet into the crop, and no force is necessary. The dose is discharged into the crop and the needle withdrawn. Doses are given twice daily before feeding so that the crop is empty, and the drug passes on quickly into the gizzard and intestine. The first dose is given about 6 hours after inoculation and further doses every night and morning for the next 3 days. One group of chicks is always kept untreated as a control. Blood smears are taken from every bird on the second and fourth days after inoculation. Curd, Davey, and Rose (1945) give the following time-table for the test:

Monday. 9 a.m. The inoculum is prepared and injected into the chicks, which are weighed and numbered and arranged in groups.
4 p.m. The first dose of drug is given. The dose is calculated for a 50-gm. chick and is dissolved or dispersed in 1 ml. of water and administered by catheter.

Tuesday. Dose repeated at 9 a.m. and 5 p.m.

Wednesday. Dose given at 9 a.m.; blood smears made at 10 a.m.; dose repeated at 5 p.m.

Thursday. Dose repeated at 9 a.m. and 5 p.m.

Friday. 9 a.m. Smears are made from the control birds, stained and examined. If the parasite level indicates that the peak of infection has been reached, smears are made from all the birds, and the test ends.

If the infection has been unexpectedly long in developing the chicks may be kept for an extra day before reading the results.

Assessment of results. The stained smears are examined, and the number of erythrocytes containing parasites in a random sample of 500 is counted. In a test performed as above the number of parasitized cells in control birds is found to be about 300. A reduction of 90 or more in the count for treated birds indicates a significant degree of activity. Curd, Davey, and Rose found that when doses were given at a level just large enough to show activity the responses of the individual chicks varied a good deal. (This is true of most chemotherapeutic

tests, and is also seen in trypanosome and leishmania infections treated with doses near the lower limit of effectiveness. It may be caused by individual differences in the rate and extent of absorption and excretion of the drug.) When the dose was large enough to suppress the infection almost completely, further increase in the size of the dose had no appreciable additional effect. For the purpose of comparison of drugs,

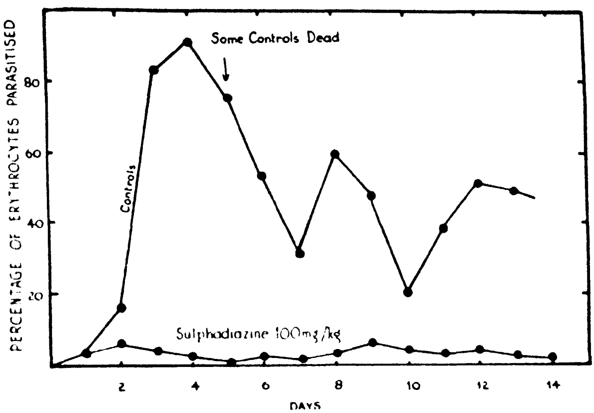


FIG. 77. The suppression of *Plasmodium gallinaceum* in the blood of the chick by sulphadiazine. Doses of 100 mg./kg. were given twice daily by mouth during the first four days of a blood-induced infection.

activity was therefore expressed as the range between doses causing minimal and maximal effects. Accurate calculations may be made by the usual methods of plotting the dose-response curves, but Curd, Davey, and Rose take the realistic view that these are a waste of time when drugs are being selected for trial against the different species of *Plasmodium* which infect man.

Fig. 77 shows the way in which *P. gallinaceum* infections are suppressed by treatment with sulphadiazine.

Sporozoite-induced Infections of *P. gallinaceum* in the Chic

Breeding of mosquitoes. For the successful breeding of mosquitoes an insectarium in which the temperature and humidity of the air can be controlled is necessary (Bertram and Gordon, 1939). It is possible to raise mosquitoes in small numbers in an incubator at 28° C. containing wet lint to keep the relative

humidity at above 70 per cent., but this is not so satisfactory as a small room set aside for the purpose.

The best species of mosquito to use for the transmission of *P. gallinaceum* is *Aëdes aegypti*. The insects are kept in a box made of fine netting stretched upon a frame. The eggs of the mosquitoes are laid singly on damp blotting-paper placed in the mosquito boxes. The blotting-paper together with the eggs may then be removed and left to dry. The eggs withstand desiccation for about 3 weeks. They may be placed when desired in large dishes of water containing a little larval medium, consisting of dog biscuit, 2 parts, and dried yeast, 1 part, ground very finely and allowed to stand in winchesters of tap-water for 2 days to acquire a rich bacterial flora on which the larvae can feed. They usually hatch overnight. Fresh larval medium should be supplied daily. The medium should not be more than 3 days old, otherwise a scum may develop on the surface and prevent larval respiration. The larvae pupate in about a week to 10 days, when they may be picked out with a pipette and placed in small dishes of water which are then put in the mosquito boxes. The adult mosquitoes emerge 2–3 days later. The mosquitoes are fed on glucose solution until they are required to lay eggs. They are then starved for 24 hours and a guinea-pig or rabbit, anaesthetized with pentobarbitone, is placed in the cage. The mosquitoes feed on the blood and will then lay eggs.

Infection of mosquitoes. Female mosquitoes, which have emerged from pupae at least 5 days previously, and have been starved for at least 24 hours, are removed from the breeding-cage with a glass tube attached to a water-pump or bellows. The insects are trapped by a piece of netting stretched across the proximal end of the tube. A chick is selected which shows a high proportion of gametocytes in the peripheral blood. Cantrell and Jordan (1946) found that the blood was most infective to mosquitoes on the day before the peak of parasitaemia was reached.

The bird is immobilized by stretching it upon a triangle of stout wire, and is introduced into a darkened cage containing the hungry mosquitoes. The insects feed readily, and rapidly become distended with blood. On two occasions during the following week the insects are fed upon guinea-pig. By the

fifth day dissection of the gut will reveal well-developed oocysts, and from the seventh day sporozoites are found in the salivary gland. To detect infection of the salivary glands, a mosquito is killed with ether, moistened with alcohol, washed with water, and placed in a drop of saline on a slide. The wings and legs are removed. With a pair of needles under the dissecting microscope, the head is gently separated from the thorax, and the salivary glands come away attached to the head. The glands are smeared on a slide, and fixed and stained with Giesma in order to demonstrate the sporozoites.

Mosquitoes remain infected for about 3 months. They should be given no further blood meals, but should be allowed access to 5 per cent. glucose solution, or slices of apple or stewed dates.

Inoculation of chicks. For the purpose of preparing an inoculum it is not necessary to separate the salivary glands from the mosquitoes. Insects fed about 12 days previously on an infected chick are taken, and the heads and thoraces ground up in a small sterile mortar with a mixture of Ringer's solution and chicken serum. Not less than 25 per cent. of serum should be used, and the final concentration of the suspension should be one mosquito in 0·2 ml. The larger fragments of mosquito are removed by straining or by light centrifuging. The equivalent of one mosquito is injected subcutaneously or intravenously into each chick.

Treatment with drug. For the examination of a drug as a causal prophylactic, it is best to administer the first dose about an hour before the inoculation of the parasites. Subsequent doses are given in the same manner as for the test upon blood-induced infections. In untreated control birds it takes from 4 to 7 days before parasites become detectable in the peripheral blood. Exoerythrocytic parasites make their appearance at the same time, and birds frequently die from blockage of brain capillaries at about the ninth to twelfth day. The birds should be examined for parasites for at least 16 days. Activity of a drug is shown by prolongation of the period during which the birds remain free from parasites in the peripheral blood, when compared with the group of controls. A dose-level of 50 mg. per kg. of sulphadiazine will usually give a significant effect. 250 mg. per kg. gives protection for at least 30 days.

Tests using Canaries infected with *P. relictum* or *P. cathemerium*

The original canary test devised by Roehl (1926) was performed by inoculating canaries with infected blood, treating with daily doses of drug for a period of 6 days, and noting the dose which caused a delay in the appearance of parasites in the peripheral blood. Buttle, Henry, and Trevan (1934) adapted the test so that a quantitative response was given by the length of the delay in appearance of parasites. In recent work the proportion of parasited erythrocytes has been used as an index of the effect of the drug, as in the chick test.

Inoculation of canaries. The infection in the canary, like that in the chick, rises to a peak during the first week and then declines. Canaries very rarely die at the peak of infection, which is much less intense than that of *P. gallinaceum* in the chick. A bird near the peak of infection is taken and blood removed from the leg or wing vein, or the bird is killed and bled from the heart. At least 5 per cent. of the cells should be parasited. The blood is prevented from clotting with heparin or sodium citrate, and is inoculated into the canaries for the test. It is customary to use female canaries, as these are less expensive than the singing males. If the injection is made into the breast muscles, parasites first appear in the peripheral blood in 4 or 5 days. If the injection is made into the leg vein, using a fine needle with the tip slightly curved, the infection mounts quickly to a maximum in 5 days, like that in the chick. A quantity of 0·02 ml. or 0·05 ml. of heparinized blood is sufficient for the inoculum.

Administration of drugs. Doses of drugs may be given twice daily as in the chick test. A fine gum-elastic or rubber catheter attached to a syringe is used, and doses of solutions or suspensions of drugs of 0·25–0·5 ml. given. Alternatively, the drug may be massed with gum acacia and syrup of glucose, and made into pellets which are dropped into the back of the pharynx. It is possible to coax canaries to eat drug-diet, but they are much more particular about their food than chicks, and a definite schedule of doses is preferable.

Assessment of effect. At the end of 6 days of treatment, blood smears are made from all of the birds, and the proportion of parasited cells present in each is determined. The effect of the

drug is shown, as in the chick test, by comparison with the infection in untreated control birds.

Plasmodium Berghei

A most useful addition to experimental malaria infections has recently been made by Vincke and Lips (1948) who discovered a parasite which they named *P. berghei* in the red cells of a rodent in the Belgian Congo. The organism grows well in laboratory mice, and kills them in 10 to 14 days after the intraperitoneal injection of infected blood. It may also be passaged in rats and hamsters. The only mammalian malarial parasites hitherto available for laboratory tests have been *P. knowlesi* and *P. cynomolgi* of monkeys; the use of mice will be much cheaper and simpler.

The assay of antimalarial drugs may be performed in a manner

Table LXXXIV. *The effect of drugs upon* Plasmodium berghei

Drug	*Dose (mg./kg.)*	*Mean percentage of red cells parasitized*	*Quinine equivalent*
Quinine	40	3	1
	23	10	
	14	24	
	8	44	
Mepacrine	10	0·5	8
	5·8	1·4	
	3·4	6·0	
	2	22·0	
Chloroquine	5	0	13
	2·9	0·9	
	1·7	4·0	
	1	38·0	
Pamaquin	5	41·0	Dose levels too low
	2·9	44·0	
	1·7	45·0	
	1	42·0	
Proguanil	10	0·1	5
	5·8	0·8	
	3·4	22·0	
	2	42·0	
Controls	..	15·0	..

very similar to that already described for *P. gallinaceum*. The blood of an infected mouse is withdrawn from the heart and diluted with heparinized saline so that an inoculum of 0·2 ml. contains about 5 million parasitized red cells. The mice to be infected are inoculated intraperitoneally with the suspension, and treatment with drugs is given twice daily as in the chick test. Doses are administered directly into the stomach, using a syringe fitted with a piece of ureteric catheter, or a blunted serum needle. Blood smears are prepared and examined on the third day, and on the day after the end of treatment. A group of infected controls are kept without drug treatment for comparison.

Table LXXXIV shows the results of a typical experiment using groups of 7 mice at each dose level. The relative activities of the drugs are different from those obtained with assays against *P. gallinaceum*.

REFERENCES

Bertram and Gordon, 1939, *Ann. Trop. Med. Parasitol.* **33**, 279.

Burn, 1948, *The Background of Therapeutics*. Oxford Medical Publications.

Buttle, Henry, and Trevan, 1934, *Biochem. J.* **28**, 426.

Cantrell and Jordan, 1946, *J. Infect. Diseases*, **78**, 153.

Curd, Davey, and Rose, 1945, *Ann. Trop. Med. Parasitol.* **39**, 139 (et seq.).

Hewitt, 1940, *Bird Malaria*. Baltimore, The Johns Hopkins Press.

Huff and Coulston, 1944, *J. Infect. Diseases*, **75**, 231.

James and Tate, 1938, *Parasitology*, **30**, 128.

Marshall, Litchfield, and White, 1942, *J. Pharmacol.* **75**, 89.

Roehl, 1926, *Beih. Arch. Schiffs- u. Tropenhyg.* **30**, 311.

Vincke and Lips, 1948, *Ann. Soc. Belge. Med. Trop.* **28**, 97.

Wiselogle, 1946, *A Survey of Antimalarial Drugs*, vol. i. Ann Arbor, Michigan, J. W. Edwards.

APPENDIX I

INTERNATIONAL BIOLOGICAL STANDARDS

Standard preparation	*Adopted*	*International unit, mg.*	*International centre*
Diphtheria Antitoxin	1922	0·0628	The State Serum Institute, Copenhagen, Denmark
Tetanus Antitoxin	1928	0·1547	
Anti-dysentery Serum (*Shiga*)	1928	0·0500	
Staphylococcus α Antitoxin	1934	0·5000	
Anti-pneumococcus Serum (Type I)	1934	0·0886	
Anti-pneumococcus Serum (Type II)	1934	0·0894	
Gas-gangrene Antitoxin (*Perfringens*)	1931	0·2660	
Gas-gangrene Antitoxin (*Vibrion septique*)	1934	0·2377	
Gas-gangrene Antitoxin (*Œdematiens*)	1934	0·2681	
Gas-gangrene Antitoxin (*Histolyticus*)	1935	0·3575	
Gas-gangrene Antitoxin (*Sordelli*)	1938	0·1334	
Old Tuberculin	1931	..	
Diphtheria Antitoxin for Flocculation Test	1935	..	
Vitamin A (Mixed Carotenes)	1931	0·001	The National Institute for Medical Research, Mill Hill, London
(Pure β Carotene)	1934	0·0006	
Vitamin B_1 (Adsorption Product of Vitamin B_1)	1931	10·0	
(Pure Synthetic Vitamin B_1)	1938	0·003	
Vitamin C (*l*-Ascorbic Acid)	1934	0·05	
Vitamin D (Irradiated Ergosterol Solution)	1931	0·1	
(Calciferol)	1934	0·000025	
Vitamin E (α-Tocopheryl Acetate)	1941	1·0	
Arsphenamine	1925	..	
Neoarsphenamine	1925	..	
Sulpharsphenamine	1925	..	
Insulin (Crude Dry Insulin Hydrochloride)	1925	0·125	
(Pure Crystalline Insulin)	1935	0·0455	
Pituitary (Posterior Lobe) Powder	1925	0·5	
Digitalis	1925	80·0	
Ouabain	1928	..	
Oestrus-producing Hormones:			
(1) Oestrone	1932	0·0001	
(2) Oestradiol Monobenzoate	1935	0·0001	
Androsterone (for Male Hormone)	1935	0·1	
Corpus Luteum Hormone (Progesterone)	1935	1·0	
Chorionic Gonadotrophin	1938	0·1	
Serum Gonadotrophin	1938	0·25	
Thyrotrophin	1938	..	
Prolactin (Galactin or Mammotrophin)	1938	0·1	
Heparin	1942	0·0077	
Penicillin	1944	0·0006	

APPENDIX II

DISTRIBUTION OF t

Degrees of freedom	*Probability* 0·50	0·10	0·05	0·01
1	1·00	6·31	12·7	63·7
2	0·82	2·92	4·30	9·92
3	0·76	2·35	3·18	5·84
4	0·74	2·13	2·78	4·60
5	0·73	2·02	2·57	4·03
6	0·72	1·94	2·45	3·71
7	0·71	1·90	2·36	3·50
8	0·71	1·86	2·31	3·36
9	0·70	1·83	2·26	3·25
10	0·70	1·81	2·23	3·17
12	0·70	1·78	2·18	3·06
14	0·69	1·76	2·14	2·98
16	0·69	1·75	2·12	2·92
18	0·69	1·73	2·10	2·88
20	0·69	1·72	2·09	2·84
22	0·69	1·72	2·07	2·82
24	0·68	1·71	2·06	2·80
26	0·68	1·71	2·06	2·78
28	0·68	1·70	2·05	2·76
30	0·68	1·70	2·04	2·75
Limit	0·674	1·645	1·960	2·576

This table, an abridgement of Table III in *Statistical Tables for Biological, Agricultural and Medical Research* by R. A. Fisher and F. Yates (Oliver & Boyd, Edinburgh), appears by permission of the authors and publishers.

APPENDIX III

DISTRIBUTION OF VARIANCE RATIO

(Values for probability of 0·05)

n_2 \ n_1	1	2	3	4	6	12	24	*Limit*
1	161	200	216	225	234	244	249	254
2	19	19	19	19	19	19	19	19
3	10·1	9·6	9·3	9·1	8·9	8·7	8·6	8·5
4	7·7	6·9	6·6	6·4	6·2	5·9	5·8	5·6
5	6·6	5·8	5·4	5·2	5·0	4·7	4·5	4·4
6	6·0	5·1	4·8	4·5	4·3	4·0	3·8	3·7
7	5·6	4·7	4·3	4·1	3·9	3·6	3·4	3·2
8	5·3	4·5	4·1	3·8	3·6	3·3	3·1	2·9
9	5·1	4·3	3·9	3·6	3·4	3·1	2·9	2·7
10	5·0	4·1	3·7	3·5	3·2	2·9	2·7	2·5
15	4·5	3·7	3·3	3·1	2·8	2·5	2·3	2·1
20	4·4	3·5	3·1	2·9	2·6	2·3	2·1	1·8
30	4·2	3·3	2·9	2·7	2·4	2·1	1·9	1·6
60	4·0	3·2	2·8	2·5	2·3	1·9	1·7	1·4
Limit	3·8	3·0	2·6	2·4	2·1	1·8	1·5	1·0

n_1 is no. of degrees of freedom for mean square to be compared with 'error'.
n_2 is no. of degrees of freedom for error mean square.

This table, an abridgement of Table V in *Statistical Tables for Biological, Agricultural and Medical Research* by R. A. Fisher and F. Yates (Oliver & Boyd, Edinburgh), appears by permission of the authors and publishers.

APPENDIX IV

PROBABILITIES OF DEATH ACCORDING TO BINOMIAL DISTRIBUTION

(a) *Tests on 5 animals*

No. of animals which die	*True mortality per cent.* 2·5	5	7·5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	92·5	95	97·5
0	·8811	·7738	·6772	·5905	·4437	·3277	·2373	·1681	·1160	·0778	·0503	·0312	·0185	·0102	·0053	·0024	·0010	·0003	·0001	·0000	·0000	..	..
1 or less	·9941	·9774	·9517	·9185	·8352	·7373	·6328	·5282	·4284	·3370	·2562	·1875	·1312	·0870	·0540	·0308	·0156	·0067	·0022	·0005	·0001	·0000	·0000
2 "	·9998	·9988	·9962	·9914	·9734	·9421	·8965	·8369	·7648	·6826	·5931	·5000	·4069	·3174	·2352	·1631	·1035	·0579	·0266	·0086	·0038	·0012	·0002
3 "	1·0000	1·0000	·9999	·9995	·9978	·9933	·9844	·9692	·9460	·9130	·8688	·8125	·7438	·6630	·5716	·4718	·3672	·2627	·1648	·0815	·0483	·0226	·0059
4 "	..	..	1·0000	1·0000	·9999	·9997	·9990	·9976	·9947	·9898	·9815	·9688	·9497	·9222	·8840	·8319	·7627	·6723	·5563	·4095	·3228	·2262	·1189

(b) *Tests on 10 animals*

No. of animals which die	*True mortality per cent.* 2·5	5	7·5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	92·5	95	97·5
0	·7763	·5987	·4586	·3487	·1969	·1074	·0563	·0282	·0135	·0060	·0025	·0010	·0003	·0001	·0000	·0000	..	..	..	..	..	..	..
1 or less	·9754	·9139	·8304	·7361	·5443	·3758	·2440	·1493	·0860	·0464	·0233	·0107	·0045	·0017	·0005	·0001	·0000	·0000	..	..	..	..	..
2 "	·9984	·9885	·9661	·9298	·8202	·6778	·5256	·3828	·2616	·1673	·0996	·0547	·0274	·0123	·0048	·0016	·0004	·0001	·0000	..	..	..	..
3 "	·9999	·9990	·9954	·9872	·9500	·8791	·7759	·6496	·5138	·3823	·2660	·1719	·1020	·0548	·0260	·0106	·0035	·0009	·0001	·0000	..	..	..
4 "	1·0000	·9999	·9996	·9984	·9901	·9672	·9219	·8497	·7515	·6331	·5044	·3770	·2616	·1662	·0949	·0473	·0197	·0064	·0014	·0001	·0000	·0000	..
5 "	..	1·0000	1·0000	·9999	·9986	·9936	·9803	·9527	·9051	·8338	·7384	·6230	·4956	·3669	·2485	·1503	·0781	·0328	·0099	·0016	·0004	·0001	·0000
6 "	..	..	..	1·0000	·9999	·9991	·9965	·9894	·9740	·9452	·8980	·8281	·7340	·6177	·4862	·3504	·2241	·1209	·0500	·0128	·0046	·0010	·0001
7 "	..	..	..	..	1·0000	·9999	·9996	·9984	·9952	·9877	·9726	·9453	·9004	·8327	·7384	·6172	·4744	·3222	·1798	·0702	·0339	·0115	·0016
8 "	..	..	..	..	..	1·0000	1·0000	·9999	·9995	·9983	·9955	·9893	·9767	·9536	·9140	·8507	·7560	·6242	·4557	·2639	·1696	·0861	·0246
9 "	..	..	..	..	..	..	..	1·0000	1·0000	·9999	·9997	·9990	·9975	·9940	·9865	·9718	·9437	·8926	·8031	·6513	·5414	·4013	·2237

(c) Tests on 15 animals

No. of animals which die	True mortality per cent.																					
	2·5	5	7·5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	92·5	95	97·5
0	·6840	·4633	·3105	·2059	·0874	·0352	·0134	·0047	·0016	·0005	·0001	·0000	·0000	..	..	..	..	..	..	..	..	..	..
1 or less	·9471	·8290	·6882	·5490	·3186	·1671	·0802	·0353	·0142	·0052	·0017	·0005	·0001	·0000	·0000	..	..	..	..	..	..	..	..
2 ,,	·9943	·9638	·9026	·8159	·6042	·3980	·2361	·1268	·0617	·0271	·0107	·0037	·0011	·0003	·0001	·0000	..	..	..	..	..	..	..
3 ,,	·9996	·9945	·9779	·9444	·8227	·6482	·4613	·2969	·1727	·0905	·0424	·0176	·0063	·0019	·0005	·0001	·0000	..	..	..	..	..	..
4 ,,	1·0000	·9994	·9962	·9873	·9383	·8358	·6865	·5155	·3519	·2173	·1204	·0592	·0255	·0093	·0028	·0007	·0001	·0000	..	..	..	..	..
5 ,,	..	·9999	·9995	·9978	·9832	·9389	·8516	·7216	·5643	·4032	·2608	·1509	·0769	·0338	·0124	·0037	·0008	·0001	·0000	..	..	..	..
6 ,,	..	1·0000	·9999	·9997	·9964	·9819	·9434	·8689	·7548	·6098	·4522	·3036	·1818	·0950	·0422	·0152	·0042	·0008	·0001	..	..	..	..
7 ,,	..	..	1·0000	1·0000	·9994	·9958	·9827	·9500	·8868	·7869	·6535	·5000	·3465	·2131	·1132	·0500	·0173	·0042	·0006	·0000	·0000	..	..
8 ,,	..	..	..	..	·9999	·9992	·9958	·9848	·9578	·9050	·8182	·6964	·5478	·3902	·2452	·1311	·0566	·0181	·0036	·0003	·0001	·0000	..
9 ,,	..	..	..	..	1·0000	·9999	·9992	·9963	·9876	·9662	·9231	·8491	·7392	·5968	·4357	·2784	·1484	·0611	·0168	·0022	·0005	·0001	..
10 ,,	..	..	..	..	..	1·0000	·9999	·9993	·9972	·9907	·9745	·9408	·8796	·7827	·6481	·4845	·3135	·1642	·0617	·0127	·0038	·0006	·0000
11 ,,	..	..	..	..	..	..	1·0000	·9999	·9995	·9981	·9937	·9824	·9576	·9095	·8273	·7031	·5387	·3518	·1773	·0556	·0221	·0055	·0004
12 ,,	..	..	..	..	..	..	..	1·0000	·9999	·9997	·9989	·9963	·9893	·9729	·9383	·8732	·7639	·6020	·3958	·1841	·0974	·0362	·0057
13 ,,	..	..	..	..	..	..	..	..	1·0000	1·0000	·9999	·9995	·9983	·9948	·9858	·9647	·9198	·8329	·6814	·4510	·3118	·1710	·0529
14 ,,	..	..	..	..	..	..	..	..	..	..	1·0000	1·0000	·9999	·9995	·9984	·9953	·9866	·9648	·9126	·7941	·6895	·5367	·3160

(*d*) *Tests on 20 animals*

No. of animals which die	*True mortality per cent.* 2·5	5	7·5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	92·5	95	97·5
0	·6027	·3585	·2103	·1216	·0388	·0115	·0032	·0008	·0002	·0000	·0000	..	..	..	..	..	..	..	..	..	..	..	..
1 or less	·9118	·7358	·5514	·3917	·1756	·0692	·0243	·0076	·0021	·0005	·0001	·0000	..	..	..	..	..	..	..	..	..	..	..
2 ,,	·9870	·9245	·8140	·6769	·4049	·2061	·0913	·0355	·0121	·0036	·0009	·0002	·0000	..	..	..	..	..	..	..	..	..	..
3 ,,	·9986	·9841	·9418	·8670	·6477	·4114	·2252	·1071	·0441	·0160	·0049	·0013	·0003	·0000	..	..	..	..	..	..	..	..	..
4 ,,	·9999	·9974	·9858	·9568	·8298	·6296	·4148	·2375	·1182	·0510	·0189	·0059	·0015	·0003	·0000	..	..	..	..	..	..	..	..
5 ,,	1·0000	·9997	·9973	·9887	·9327	·8042	·6172	·4164	·2454	·1256	·0553	·0207	·0064	·0016	·0003	·0000	..	..	..	..	..	..	..
6 ,,	..	1·0000	·9996	·9976	·9781	·9133	·7858	·6080	·4166	·2500	·1299	·0577	·0214	·0065	·0015	·0003	·0000	..	..	..	..	..	..
7 ,,	..	..	·9999	·9996	·9941	·9679	·8982	·7723	·6010	·4159	·2520	·1316	·0580	·0210	·0060	·0013	·0002	·0000	..	..	..	..	..
8 ,,	..	..	1·0000	·9999	·9987	·9900	·9591	·8867	·7624	·5956	·4143	·2517	·1308	·0565	·0196	·0051	·0009	·0001	..	..	..	..	..
9 ,,	..	..	..	1·0000	·9998	·9974	·9861	·9520	·8782	·7553	·5911	·4119	·2493	·1275	·0532	·0171	·0039	·0006	·0000	..	..	..	..
10 ,,	..	..	..	..	1·0000	·9994	·9961	·9829	·9468	·8725	·7507	·5881	·4086	·2447	·1218	·0480	·0139	·0026	·0002	·0000	..	..	..
11 ,,	..	..	..	..	..	·9999	·9991	·9949	·9804	·9435	·8692	·7483	·5857	·4044	·2376	·1133	·0409	·0100	·0013	·0001	·0000	..	..
12 ,,	..	..	..	..	..	1·0000	·9998	·9987	·9940	·9790	·9420	·8684	·7180	·5841	·3990	·2277	·1018	·0321	·0059	·0004	·0001	..	..
13 ,,	..	..	..	..	..	..	1·0000	·9997	·9985	·9935	·9786	·9423	·8701	·7500	·5834	·3920	·2142	·0867	·0219	·0024	·0004	·0000	..
14 ,,	..	..	..	..	..	..	..	1·0000	·9997	·9984	·9936	·9793	·9447	·8744	·7546	·5836	·3828	·1958	·0673	·0113	·0027	·0003	·0000
15 ,,	..	..	..	..	..	..	..	..	1·0000	·9997	·9985	·9941	·9811	·9490	·8818	·7625	·5852	·3704	·1702	·0432	·0142	·0026	·0001
16 ,,	..	..	..	..	..	..	..	..	..	1·0000	·9997	·9987	·9951	·9840	·9556	·8929	·7748	·5886	·3523	·1330	·0582	·0159	·0014
17 ,,	..	..	..	..	..	..	..	..	..	..	1·0000	·9998	·9991	·9964	·9879	·9645	·9087	·7939	·5951	·3231	·1860	·0755	·0130
18 ,,	..	..	..	..	..	..	..	..	..	..	..	1·0000	·9999	·9995	·9979	·9924	·9757	·9308	·8244	·6083	·4486	·2642	·0882
19 ,,	..	..	..	..	..	..	..	..	..	..	..	..	1·0000	1·0000	·9998	·9992	·9968	·9885	·9612	·8784	·7897	·6415	·3973

(e) Tests on 30 animals

No. of animals which die	True mortality per cent.																					
	2·5	*5*	*7·5*	*10*	*15*	*20*	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*	*70*	*75*	*80*	*85*	*90*	*92·5*	*95*	*97·5*
0	·4679	·2146	·0964	·0424	·0076	·0012	·0002	·0000	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
1 or less	·8278	·5535	·3310	·1837	·0480	·0105	·0020	·0003	·0000	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2 ,,	·9616	·8122	·6068	·4114	·1514	·0442	·0106	·0021	·0003	·0000	..	..	..	..	..	..	..	..	..	..	..	..	..
3 ,,	·9936	·9392	·8155	·6474	·3217	·1227	·0374	·0093	·0019	·0003	·0000	..	..	..	..	..	..	..	..	..	..	..	..
4 ,,	·9992	·9844	·9297	·8245	·5245	·2552	·0979	·0302	·0075	·0015	·0002	·0000	..	..	..	..	..	..	..	..	..	..	..
5 ,,	·9999	·9967	·9779	·9268	·7106	·4275	·2026	·0766	·0233	·0057	·0011	·0002	·0000	..	..	..	..	..	..	..	..	..	..
6 ,,	1·0000	·9994	·9942	·9742	·8474	·6070	·3481	·1595	·0586	·0172	·0040	·0007	·0001	..	..	..	..	..	..	..	..	..	..
7 ,,	..	·9999	·9987	·9922	·9302	·7608	·5143	·2814	·1238	·0435	·0121	·0026	·0004	·0000	..	..	..	..	..	..	..	..	..
8 ,,	..	1·0000	·9997	·9980	·9722	·8713	·6736	·4315	·2247	·0940	·0312	·0081	·0016	·0002	·0000	..	..	..	..	..	..	..	..
9 ,,	..	..	1·0000	·9995	·9903	·9389	·8034	·5888	·3575	·1763	·0694	·0214	·0050	·0009	·0001	..	..	..	..	..	..	..	..
10 ,,	..	..	..	·9999	·9971	·9744	·8943	·7304	·5078	·2915	·1350	·0494	·0138	·0029	·0004	·0000	..	..	..	..	..	..	..
11 ,,	..	..	..	1·0000	·9992	·9905	·9493	·8407	·6548	·4311	·2327	·1002	·0334	·0083	·0014	·0002	·0000	..	..	..	..	..	..
12 ,,	..	..	..	..	·9998	·9969	·9784	·9155	·7802	·5785	·3592	·1808	·0714	·0212	·0045	·0006	·0001	..	..	..	..	..	..
13 ,,	..	..	..	..	1·0000	·9991	·9918	·9599	·8737	·7145	·5025	·2923	·1356	·0481	·0124	·0021	·0002	·0000	..	..	..	..	..
14 ,,	..	..	..	..	..	·9998	·9973	·9831	·9348	·8246	·6448	·4278	·2309	·0971	·0301	·0064	·0008	·0001	..	..	..	..	..
15 ,,	..	..	..	..	..	·9999	·9992	·9936	·9699	·9029	·7691	·5722	·3552	·1754	·0652	·0169	·0027	·0002	..	..	..	..	..
16 ,,	..	..	..	..	..	1·0000	·9998	·9979	·9876	·9519	·8644	·7077	·4975	·2855	·1263	·0401	·0082	·0009	·0000	..	..	..	..
17 ,,	..	..	..	..	..	..	·9999	·9994	·9955	·9788	·9286	·8192	·6408	·4215	·2198	·0845	·0216	·0031	·0002	..	..	..	..
18 ,,	..	..	..	..	..	..	1·0000	·9998	·9986	·9917	·9666	·8998	·7673	·5689	·3452	·1593	·0507	·0095	·0008	·0000	..	..	..
19 ,,	..	..	..	..	..	..	..	1·0000	·9996	·9971	·9862	·9506	·8650	·7085	·4922	·2696	·1057	·0256	·0029	·0001	..	..	..
20 ,,	..	..	..	..	..	..	..	..	·9999	·9991	·9950	·9786	·9306	·8237	·6425	·4112	·1966	·0611	·0097	·0005	·0000	..	..
21 ,,	..	..	..	..	..	..	..	..	1·0000	·9998	·9984	·9919	·9688	·9060	·7753	·5684	·3264	·1287	·0278	·0020	·0003	·0000	..
22 ,,	..	..	..	..	..	..	..	..	..	1·0000	·9996	·9974	·9879	·9565	·8742	·7186	·4857	·2392	·0698	·0078	·0013	·0001	..
23 ,,	..	..	..	..	..	..	..	..	..	..	·9999	·9993	·9960	·9828	·9414	·8405	·6519	·3930	·1526	·0258	·0058	·0006	·0000
24 ,,	..	..	..	..	..	..	..	..	..	..	1·0000	·9998	·9989	·9943	·9767	·9234	·7974	·5725	·2894	·0732	·0221	·0033	·0001
25 ,,	..	..	..	..	..	..	..	..	..	..	..	1·0000	·9998	·9985	·9925	·9698	·9021	·7448	·4755	·1755	·0703	·0156	·0008
26 ,,	..	..	..	..	..	..	..	..	..	..	..	..	1·0000	·9997	·9981	·9907	·9626	·8773	·6783	·3526	·1845	·0608	·0064
27 ,,	..	..	..	..	..	..	..	..	..	..	..	..	..	1·0000	·9997	·9979	·9894	·9558	·8486	·5886	·3932	·1878	·0384
28 ,,	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1·0000	·9997	·9980	·9895	·9520	·8163	·6690	·4465	·1722
29 ,,	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1·0000	·9998	·9988	·9924	·9576	·9036	·7854	·5321

APPENDIX V

DISTRIBUTION OF χ^2

Degrees of freedom	*Probability*					
	0·99	*0·95*	*0·50*	*0·10*	*0·05*	*0·01*
1	0·0002	0·004	0·5	2·7	3·8	6·6
2	0·02	0·1	1·4	4·6	6·0	9·2
3	0·1	0·4	2·4	6·3	7·8	11·3
4	0·3	0·7	3·4	7·8	9·5	13·3
5	0·6	1·1	4·4	9·2	11·1	15·1
6	0·9	1·6	5·3	10·6	12·6	16·8
7	1·2	2·2	6·3	12·0	14·1	18·5
8	1·6	2·7	7·3	13·4	15·5	20·1
9	2·1	3·3	8·3	14·7	16·9	21·7
10	2·6	3·9	9·3	16·0	18·3	23·2
12	3·6	5·2	11·3	18·5	21·0	26·2
14	4·7	6·6	13·3	21·1	23·7	29·1
16	5·8	8·0	15·3	23·5	26·3	32·0
18	7·0	9·4	17·3	26·0	28·9	34·8
20	8·3	10·9	19·3	28·4	31·4	37·6
22	9·5	12·3	21·3	30·8	33·9	40·3
24	10·9	13·8	23·3	33·2	36·4	43·0
26	12·2	15·4	25·3	35·6	38·9	45·6
28	13·6	16·9	27·3	37·9	41·3	48·3
30	15·0	18·5	29·3	40·3	43·8	50·9
	When χ^2 has more than 30 degrees of freedom, calculate $\sqrt{(2\chi^2)} - \sqrt{(2f-1)}$, where f is the number of degrees of freedom, and compare with the following values:					
f	−2·33	−1·64	0·00	1·28	1·64	2·33

This table, an abridgement of Table IV in *Statistical Tables for Biological, Agricultural and Medical Research* by R. A. Fisher and F. Yates (Oliver & Boyd, Edinburgh), appears by permission of the authors and publishers.

APPENDIX VI

TRANSFORMATION OF PERCENTAGES TO PROBITS

%	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	..	2·67	2·95	3·12	3·25	3·36	3·45	3·52	3·59	3·66
10	3·72	3·77	3·82	3·87	3·92	3·96	4·01	4·05	4·08	4·12
20	4·16	4·19	4·23	4·26	4·29	4·33	4·36	4·39	4·42	4·45
30	4·48	4·50	4·53	4·56	4·59	4·61	4·64	4·67	4·69	4·72
40	4·75	4·77	4·80	4·82	4·85	4·87	4·90	4·92	4·95	4·97
50	5·00	5·03	5·05	5·08	5·10	5·13	5·15	5·18	5·20	5·23
60	5·25	5·28	5·31	5·33	5·36	5·39	5·41	5·44	5·47	5·50
70	5·52	5·55	5·58	5·61	5·64	5·67	5·71	5·74	5·77	5·81
80	5·84	5·88	5·92	5·95	5·99	6·04	6·08	6·13	6·18	6·23
90	6·28	6·34	6·41	6·48	6·55	6·64	6·75	6·88	7·05	7·33
	0·0	*0·1*	*0·2*	*0·3*	*0·4*	*0·5*	*0·6*	*0·7*	*0·8*	*0·9*
99	7·33	7·37	7·41	7·46	7·51	7·58	7·65	7·75	7·88	8·09

This table is reproduced from Table 1 of *Probit Analysis* by D. J. Finney (Cambridge University Press, London) by permission of the publishers.

APPENDIX VII

PROBIT WEIGHTING COEFFICIENTS

Expected probit	*0·0*	*0·1*	*0·2*	*0·3*	*0·4*	*0·5*	*0·6*	*0·7*	*0·8*	*0·9*
1·0	0·001	0·001	0·001	0·002	0·002	0·003	0·005	0·006	0·008	0·011
2·0	0·015	0·019	0·025	0·031	0·040	0·050	0·062	0·076	0·092	0·110
3·0	0·131	0·154	0·180	0·208	0·238	0·269	0·302	0·336	0·370	0·405
4·0	0·439	0·471	0·503	0·532	0·558	0·581	0·601	0·616	0·627	0·634
5·0	0·637	0·634	0·627	0·616	0·601	0·581	0·558	0·532	0·503	0·471
6·0	0·439	0·405	0·370	0·336	0·302	0·269	0·238	0·208	0·180	0·154
7·0	0·131	0·110	0·092	0·076	0·062	0·050	0·040	0·031	0·025	0·019
8·0	0·015	0·011	0·008	0·006	0·005	0·003	0·002	0·002	0·001	0·001

This table is reproduced from Table 4 of *Probit Analysis* by D. J. Finney (Cambridge University Press, London) by permission of the publishers.

APPENDIX VIII. TABLE OF WORKING PROBITS

	Percentage kill																				
Y	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100
2·0	1·70	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·1	1·79	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·2	1·88	8·19	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·3	1·97	6·77	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·4	2·06	5·74	9·42	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·5	2·15	5·00	7·85	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·6	2·23	4·47	6·70	8·93	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·7	2·32	4·09	5·85	7·62	9·38	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·8	2·41	3·82	5·23	6·64	8·05	9·46	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
2·9	2·49	3·63	4·77	5·90	7·04	8·18	9·31	..	..	..	..	..	..	..	..	..	..	..	..	..	..
3·0	2·58	3·50	4·43	5·36	6·28	7·21	8·14	9·06	9·99	..	..	..	..	..	..	..	..	..	..	..	..
3·2	2·74	3·38	4·01	4·64	5·28	5·91	6·54	7·18	7·81	8·44	9·08	9·71	..	..	..	..	..	..	..	..	..
3·4	2·91	3·36	3·81	4·26	4·71	5·16	5·61	6·06	6·51	6·96	7·41	7·86	8·32	8·77	9·22	9·67	..	..	..	..	..
3·6	3·06	3·39	3·73	4·06	4·40	4·73	5·06	5·40	5·73	6·07	6·40	6·73	7·07	7·40	7·74	8·07	8·40	8·74	9·07	9·41	9·74
3·8	3·21	3·46	3·72	3·98	4·24	4·49	4·75	5·01	5·27	5·52	5·78	6·04	6·30	6·55	6·81	7·07	7·33	7·58	7·84	8·10	8·36
4·0	3·34	3·55	3·76	3·96	4·17	4·38	4·58	4·79	5·00	5·20	5·41	5·62	5·82	6·03	6·24	6·44	6·65	6·86	7·06	7·27	7·48
4·2	3·47	3·64	3·81	3·99	4·16	4·33	4·50	4·68	4·85	5·02	5·19	5·37	5·54	5·71	5·89	6·06	6·23	6·40	6·58	6·75	6·32
4·4	3·58	3·73	3·88	4·03	4·18	4·33	4·48	4·63	4·78	4·93	5·08	5·23	5·38	5·53	5·68	5·83	5·98	6·13	6·28	6·43	6·58
4·6	3·66	3·80	3·94	4·07	4·21	4·34	4·48	4·61	4·75	4·89	5·02	5·16	5·29	5·43	5·57	5·70	5·84	5·97	6·11	6·24	6·38
4·8	3·72	3·85	3·98	4·11	4·24	4·36	4·49	4·62	4·75	4·87	5·00	5·13	5·26	5·39	5·51	5·64	5·77	5·90	6·03	6·15	6·28
5·0	3·75	3·87	4·00	4·12	4·25	4·37	4·50	4·62	4·75	4·87	5·00	5·13	5·25	5·38	5·50	5·63	5·75	5·88	6·00	6·13	6·25
5·2	3·72	3·85	3·97	4·10	4·23	4·36	4·49	4·61	4·74	4·87	5·00	5·13	5·25	5·38	5·51	5·64	5·76	5·89	6·02	6·15	6·28
5·4	3·62	3·76	3·89	4·03	4·16	4·30	4·43	4·57	4·71	4·84	4·98	5·11	5·25	5·39	5·52	5·66	5·79	5·93	6·06	6·20	6·34
5·6	3·42	3·57	3·72	3·87	4·02	4·17	4·32	4·47	4·62	4·77	4·92	5·07	5·22	5·37	5·52	5·67	5·82	5·97	6·12	6·27	6·42
5·8	3·08	3·25	3·42	3·60	3·77	3·94	4·11	4·29	4·46	4·63	4·81	4·98	5·15	5·32	5·50	5·67	5·84	6·01	6·19	6·36	6·53
6·0	2·52	2·73	2·94	3·14	3·35	3·56	3·76	3·97	4·18	4·38	4·59	4·80	5·00	5·21	5·42	5·62	5·83	6·04	6·24	6·45	6·66
6·2	1·64	1·90	2·16	2·42	2·67	2·93	3·19	3·45	3·70	3·96	4·22	4·48	4·73	4·99	5·25	5·51	5·76	6·02	6·28	6·54	6·79
6·4	0·26	0·59	0·93	1·26	1·60	1·93	2·26	2·60	2·93	3·27	3·60	3·93	4·27	4·60	4·94	5·27	5·60	5·94	6·27	6·61	6·94
6·6	..	..	..	..	..	0·33	0·78	1·23	1·68	2·14	2·59	3·04	3·49	3·94	4·39	4·84	5·29	5·74	6·19	6·64	7·09
6·8	..	..	..	..	..	..	..	..	..	0·29	0·92	1·56	2·19	2·82	3·46	4·09	4·72	5·36	5·99	6·62	7·26
7·0	..	..	..	..	..	..	..	..	..	..	..	..	0·01	0·94	1·86	2·79	3·72	4·64	5·57	6·50	7·42
7·1	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0·69	1·82	2·96	4·10	5·23	6·37	7·51
7·2	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0·54	1·95	3·36	4·77	6·18	7·59
7·3	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0·62	2·38	4·15	5·91	7·68
7·4	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1·07	3·30	5·53	7·77
7·5	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	2·15	5·00	7·85
7·6	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	0·58	4·26	7·94
7·7	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	3·23	8·03
7·8	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	1·81	8·12
7·9	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	8·21

INDEX